

STATISTICAL RESEARCH REPORT
Institute of Mathematics
University of Oslo

No 3
February 1976

DEVIATIONS FROM TOTAL INFORMATION
AND FROM TOTAL IGNORANCE AS
MEASURES OF INFORMATION

by

Erik N. Torgersen

CONTENTS

1.	Introduction and acknowledgements	1.1 - 1.7
2.	Notations and a few useful tools	2.1 - 2.9
3.	Comparison of experiments w.r.t. a totally informative experiment	3.1 - 3.13
4.	Comparison of experiments w.r.t. a totally non informative experiment	4.1 - 4.6
5.	Two inequalities	5.1 - 5.11
6.	Replicated dichotomies	6.1 - 6.24
7.	Replicated experiments when the parameter set is finite	7.1 - 7.15
8.	Replicated experiments when the parameter set is countable	8.1 - 8.20
9.	Replicated translation experiments on the integers	9.1 - 9.8
10.	References	10.1 - 10.2

1. Introduction and acknowledgements.

We shall in this paper consider two measures of the contents of information in statistical experiments. They are both based on Le Cam's, [12], notion of a deficiency of one experiment w.r.t. another.

The idea is simply this. Any experiment is more informative than an experiment where the chance mechanism does not depend on which of the underlying theories are true. Let \mathcal{M}_i denote an experiment with this property, i.e. \mathcal{M}_i is a totally non informative experiment. If \mathcal{E} is any experiment then we might consider it as containing much or little information according to whether it is, respectively, far away from \mathcal{M}_i or close to \mathcal{M}_i . As a measure of this distance, the deficiency of \mathcal{M}_i w.r.t. \mathcal{E} is an obvious candidate.

Let us, before proceeding, agree - following Le Cam [12] - to use the notation $\delta(\mathcal{E}, \mathcal{F})$ for the deficiency of an experiment \mathcal{E} w.r.t. an experiment \mathcal{F} . This deficiency is defined for pairs $(\mathcal{E}, \mathcal{F})$ of experiments having the same set of underlying theories, or parameter set. We shall reserve the letter Θ as notation for the parameter set.

Thus our first proposal for a measure of the content of information in the experiment \mathcal{E} is the number $\delta(\mathcal{M}_i, \mathcal{E})$. If this distance is small - then the chance mechanism governing the random outcome is almost independent of the various explaining theories in Θ . If, on the other hand, this distance is large then there

are situations where an observation of \mathcal{E} is helpful.

We may also consider the experiment of directly observing the underlying theory θ in Θ . This experiment, which is more informative than any other experiment, will be denoted by \mathcal{U}_a . An experiment \mathcal{E} may be considered to contain much or little information according to whether \mathcal{E} is close to \mathcal{U}_a or far away from \mathcal{U}_a . Thus we arrive at the deficiency of \mathcal{E} w.r.t. \mathcal{U}_a ; i.e. $\delta(\mathcal{E}, \mathcal{U}_a)$, as a measure of the content of information in \mathcal{E} .

A small value of $\delta(\mathcal{E}, \mathcal{U}_a)$ tells that an observation of \mathcal{E} , provided it is properly used, is almost as good as knowing the unknown parameter. A large value, on the other hand, tells that there are decision problems such that any decision procedure is risky for some of the underlying theories.

The values of these deficiencies are often extremely large* for all experiments \mathcal{E} under consideration. This reflects the fact that it may be much too ambitious to compare with total information and much too modest to compare with no information.

It will be shown in section 3 that the evaluation of $\delta(\mathcal{E}, \mathcal{U}_a)$ is related to the problem of guessing the true value of θ . This may, alternatively, be viewed as a problem of finding optimal confidence regions with extreme accuracy. If we relaxed the requirement on accuracy, then we might hope to find other and more realistic measures of information than $\delta(\mathcal{E}, \mathcal{U}_a)$. Thus one might expect that the usefulness of the measure $\delta(\mathcal{E}, \mathcal{U}_a)$ is limited to situations where the space of underlying

* i.e. very close to $2^{-2[\#\Theta]^{-1}}$. Any deficiency is $\leq 2^{-2[\#\Theta]^{-1}}$.

theories is, in some sense fairly small. If \mathcal{E} is dominated then, as explained in section 3, the infimum of the probabilities of guessing the true value of θ is zero for any guessing procedure, when θ is not countable. We have, for this reason, chosen to limit this study to experiments with countable parameter sets.

Having made this limitation, we shall see that these deficiencies have interesting properties. We shall, in particular, see that $\delta(\mathcal{E}, \mathcal{M}_a)$ may be used to define a "capacity" for replicated experiments. This capacity, a non negative number * , provides an upper bound, and often the exact value, for exponential rates of convergence, as the number of replications $\rightarrow \infty$ for minimum Baye's risks and minimax risks in various decision problems.

The content is, section by section, as follows :

A few of the basic definitions and a few results from the theory of comparison of experiments are summarized in section 2. Weighted deficiencies are mentioned in Le Cam's fundamental paper [12]. We have, as these deficiencies will be found helpful, included a brief exposition of some of the elementary properties of weighted deficiencies.

It is shown in section 3 that $\frac{1}{2} \delta(\mathcal{E}, \mathcal{M}_a)$ is the minimax probability of an incorrect guess of the true value of θ . Similarly the weighted deficiencies of \mathcal{E} w.r.t. \mathcal{M}_a equals twice the minimum Baye's probability of the same event. Some

*) If $\sigma(\mathcal{E})$ is defined as in section 8 then this capacity is $-\log \sigma(\mathcal{E})$.

bounds, upper and lower, for these deficiencies in terms of deficiencies for restrictions to sub parameter sets are given. Deficiencies w.r.t. \mathcal{M}_a for testing problems are also considered. We conclude the section by computing deficiencies within a class of very peculiar experiments. In spite of their extreme inhomogeneity these experiments are useful as examples and as tools obtain results for more "normal" experiments.

The deficiency $\delta(\mathcal{M}_i, \mathcal{G})$ is, as is explained in section 4, the minimax risk for the problem of guessing the true value of θ , when no observations are available and the loss is measured by statistical distance. If we restrict attention to testing problems then the corresponding deficiency reduces to the half diameter of \mathcal{G} for statistical distance. A few simple inequalities are given and the deficiencies $\delta(\mathcal{M}_i, \mathcal{G})$ and $\delta(\mathcal{G}, \mathcal{M}_a)$ are compared for some of the particular experiments mentioned above.

The problem of consistency of the measures $\delta(\mathcal{M}_i, \mathcal{G})$ and $\delta(\mathcal{G}, \mathcal{M}_a)$ for experiments close to \mathcal{M}_a or close to \mathcal{M}_i , is investigated in section 5. If Θ is finite, then consistency follows directly from the compactness of convergence for Le Cam's distance Δ , - see [15] or [19]. We shall, however, need more precise results and these are given in section 5. If Θ is infinite then the deficiencies $\delta(\mathcal{G}, \mathcal{M}_a)$ and $\delta(\mathcal{M}_i, \mathcal{G})$ may both be close to the maximum value, $\delta(\mathcal{M}_i, \mathcal{M}_a)$. If Θ is finite then a large value of one of these quantities imply that the other is small.

The case of dichotomies, i.e. the case where Θ has two elements, is investigated in section 6. It is, by an identity, shown how the Hellinger transform may be obtained from the weighted deficiencies w.r.t. \mathcal{M}_a . Some upper - and lower bounds for the Hellinger transform in terms of weighted deficiencies w.r.t. \mathcal{M}_a are given.

The last half of section 6 and most of section 7-9 are devoted to replicated experiments. If \mathcal{E} is any experiment and n is a positive integer, then \mathcal{E}^n denotes the experiment obtained by combining n independent replications of \mathcal{E} . All limits, if not otherwise stated, are taken as $n \rightarrow \infty$.

The inequalities in section 6 imply readily that $\sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)}$ converges for any dichotomy \mathcal{E} to a number $C(\mathcal{E})$ in $[0,1]$. This may easily be extended, showing that the n -th root of minimum Baye's risk as well as the n -th root of many other functionals converges, provided some mild regularity conditions are satisfied, to the same limit $C(\mathcal{E})$. Now Chernoff, [6], proved that the n -th root of the minimum Baye's probability of guessing wrongly the underlying distribution converged to the minimum of the Hellinger transform. It follows that $C(\mathcal{E})$ and this minimum is the same number. In their paper [9], Effron and Truax established an asymptotic expansion of the minimum Baye's probability mentioned above. Extending this we derive the asymptotic expansions of many other functionals as $\delta(\cdot, \mathcal{M}_a)$ and minimum Baye's risk in various decision problems. As an application we consider the asymptotic consequences of choosing the wrong prior distribution.

As pairwise equivalence for ordered experiments imply equivalence one might hope that comparison w.r.t. \mathcal{M}_i and \mathcal{M}_a may, to some extent, be expressed in terms of dichotomies. If Θ is infinite then - as is shown in section 8 - this does not hold. If Θ is finite, however, then - as is shown in section 7 - the approximations are readily expressed in terms of dichotomies. As in section 6 we get at the same time the exponential rate of convergence of many other functionals.

Generalizing Chernoff's result in another direction we consider optimal r -point confidence sets. It turns out that the exponential rates of convergence of minimax probabilities (and of minimum Baye's probabilities) of not covering the true value has simple expressions in terms of the Hellinger transform

It is shown in section 8 that there corresponds to any experiment \mathcal{E} with countable parameter set a non negative constant , $-\log \sigma(\mathcal{E})$, which in many respects plays the role of a capacity. The problem of finding a, hopefully manageable, explicit expression for $\sigma(\mathcal{E})$ is not solved, but we do have upper and lower bounds. It may easily, happen that $\sigma(\mathcal{E}) = 1$. If so, then $\delta(\mathcal{E}^n, \mathcal{M}_a) \geq 1$ for all n , so that convergence does not occur. Thus, if convergence occurs at all, then the speed of convergence is necessarily exponential. The constant $\sigma(\mathcal{E})$ may also be related to the distribution of the required number of replicates needed in order that the optimal estimators stabilizes. Examples illustrating various possibilities are given.

If \mathcal{E} is a translation experiment on the integers then, as is shown in section 9, $\sigma(\mathcal{E}) < 1$. It follows that exponential convergence always takes place in this case. The investigation of these experiments is greatly simplified by the fact that any translation invariant maximum likelihood estimator is optimal for the guessing problem. We conclude the section - and the paper -

by an example showing that dramatic improvement may be obtained by adding a single replication. This example brings us close to another and related topic - the relative amount of information in additional observations. We refer the reader to Le Cam [14] and to the author [20] for some results in this direction.

Acknowledgements.

The author is grateful to Sigrid Cordtsen and Gerd Salter who typed this paper.

2. Notations and a few useful tools.

A few facts on comparison of experiments are summarized below. The reader having some knowledge on the subject may, after a glance at the notations, pass directly to section 3 and consult this section when needed.

Expositions of the theory of comparison of experiments may be found in Blackwell and Girshick [3], Heyer [11], Le Cam [15] and the author [23].

An experiment \mathcal{E} will in this paper be defined as a family of probability measures on a common measurable space. This measurable space and the index set of the family are called, respectively, the sample space of \mathcal{E} and the parameter set of \mathcal{E} . Thus an experiment \mathcal{E} with sample space (χ, \mathcal{A}) and parameter set Θ is a family $(P_\theta; \theta \in \Theta)$ of probability measures on (χ, \mathcal{A}) . This experiment may be denoted by $(P_\theta; \theta \in \Theta)$ or by $(\chi, \mathcal{A} ; P_\theta; \theta \in \Theta)$.

New experiments may be derived from old ones by various devices. If Θ_0 is a sub set of Θ and $\mathcal{E} = (P_\theta; \theta \in \Theta)$ is an experiment with parameter set Θ then \mathcal{E}_{Θ_0} denote the restriction $(P_\theta; \theta \in \Theta_0)$ of \mathcal{E} to Θ_0 . If $\mathcal{E}_i = (\chi_i, \mathcal{A}_i, P_{\theta,i}; \theta \in \Theta)$; $1 \leq i \leq n$ are experiments then

$(\prod_{i=1}^n (\chi_i, \mathcal{A}_i), \prod_{i=1}^n P_{\theta,i}; \theta \in \Theta)$ is called the product of $\mathcal{E}_1, \mathcal{E}_2, \dots, \dots, \mathcal{E}_n$ and it is denoted by $\mathcal{E}_1 \times \dots \times \mathcal{E}_n$ or $\prod_i \mathcal{E}_i$. If $\mathcal{E}_1 = \mathcal{E}_2 = \dots = \mathcal{E}_n = \mathcal{E}$ then we may write \mathcal{E}^n instead $\mathcal{E}_1 \times \dots \times \mathcal{E}_n$. Experiments \mathcal{E}^n ; $n=1,2,\dots$ are called replicated experiments.

Let $\mathcal{E} = (P_\theta; \theta \in \Theta)$ be an experiment. A random variable X will be called an observation of \mathcal{E} if the distribution of X

under θ is P_θ . Note that (X_1, X_2, \dots, X_n) is an observation of $\mathcal{G}_1 \times \dots \times \mathcal{G}_n$ if X_1, \dots, X_n are independent and X_i ; $i=1, \dots, n$ is an observation of \mathcal{G}_i .

We shall use the notation $\underline{\mathcal{M}}_i$ for an experiment $(P_\theta; \theta \in \Theta)$ where P_θ does not depend on θ . An experiment of this type will be called totally non informative. As any two such experiments are equivalent for Le Cam's distance Δ we do not bother to distinguish between them. The restriction of $\underline{\mathcal{M}}_i$ to a sub parameter set Θ_0 will usually also be denoted by $\underline{\mathcal{M}}_i$, instead of $[\underline{\mathcal{M}}_i]_{\Theta_0}$.

If Θ is countable (i.e. finite or enumerably infinite) then an experiment $(P_\theta; \theta \in \Theta)$ will be called totally informative provided P_{θ_1} and P_{θ_2} are mutually singular when $\theta_1 \neq \theta_2$. Any two such experiments having the same parameter set Θ are clearly equivalent. An experiment with these properties, as well as the restrictions of such experiments to sub parameter sets, will be denoted by $\underline{\mathcal{M}}_a$.

No confusion is likely to arise from our ambiguous uses of the symbols $\underline{\mathcal{M}}_i$ and $\underline{\mathcal{M}}_a$. The reason is that these symbols usually occur in expressions like $\delta(\underline{\mathcal{M}}_i, \mathcal{G})$ and $\delta(\mathcal{G}, \underline{\mathcal{M}}_a)$ and these expressions are only defined when $\underline{\mathcal{M}}_i$, respectively $\underline{\mathcal{M}}_a$, have the same parameter set as the experiment \mathcal{G} .

Important functionals of experiments may be defined as follows. Let φ be a homogenous and measurable function on $[0, \infty]^\Theta$.

If $\mathcal{G} = (P_\theta : \theta \in \Theta)$ is dominated by the σ finite measure μ then we may put $\varphi(\mathcal{G}) = \int \varphi(dP_\theta/d\sigma ; \theta \in \Theta) d\sigma$ provided this integral exists. It is easily seen that neither the existence nor the value of $\varphi(\mathcal{G})$ depends on the dominating measure σ . Instead of $\varphi(\mathcal{G})$ it is occasionally convenient to use the notation $\varphi(\mathcal{G}) = \int \varphi(dP_\theta : \theta \in \Theta)$. We mention a few functionals of this form.

If P and Q are probability measures then $\int |dP - dQ|$ is the statistical distance between them.

The affinity between P and Q is the number $\gamma(P, Q) = \int \sqrt{dP dQ}$ while the Hellinger distance between P and Q is $D(P, Q) = \sqrt{\int (\sqrt{dP} - \sqrt{dQ})^2} = \sqrt{2(1 - \gamma(P, Q))}$. This distance is equivalent to the statistical distance since, see [13],

$$D^2(P, Q) \leq \|P - Q\| \leq 2D(P, Q).$$

The Hellinger transform of $\mathcal{G} = (P_\theta : \theta \in \Theta)$ may be defined as the map $H_{\mathcal{G}}$ which to each prior distribution t with finite support associates the number $H_{\mathcal{G}}(t) = \int \prod_{\theta} dP_{\theta}^{t_{\theta}}$. Thus $\gamma(P, Q) = H_{\mathcal{G}}(\frac{1}{2}, \frac{1}{2})$ where $\mathcal{G} = (P, Q)$.

If Θ is finite and $\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_n$ have, respectively standard measures S_1, S_2, \dots, S_n , then, for each homogeneous measurable function φ on $[0, \infty]^\Theta$ which is bounded (or non negative) on Λ we have :

$$\varphi(\mathcal{G}_1 x \dots x \mathcal{G}_n) = \int \varphi(x^1 x^2 \dots x^n) (S_1 x S_2 \dots x S_n) (d(x^1, x^2, \dots, x^n))$$

where $x^1 x^2 \dots x^n = \{x_{\theta}^1 x_{\theta}^2 \dots x_{\theta}^n ; \theta \in \Theta\}$.

It follows, for a general parameter set, that the Hellinger transform converts products of experiments into products of functions i.e.:

$$H_{\mathcal{G}_1 \times \dots \times \mathcal{G}_n} = H_{\mathcal{G}_1} \cdot H_{\mathcal{G}_2} \dots H_{\mathcal{G}_n}.$$

If Θ is finite then, using the notations in [19], Γ denotes the class of sub linear functions γ on R^Θ such that

$$\gamma(e^\theta) = \gamma(-e^\theta) ; \theta \in \Theta \quad \text{and} \quad \sum_{\theta} \gamma(e^\theta) = 1.$$

Here e^θ , for each θ , is the θ -th unit vector in R^Θ , i.e. $e^\theta = (0, \dots, \overset{\theta}{1}, \dots, 0)$. The sub class of Γ consisting of those functions in Γ which are maximums of k -linear functionals will be denoted by Γ_k . A function γ will be called super linear if $-\gamma$ is sub linear.

In [12] Le Cam introduced the notion of ϵ -deficiency of one experiment relative to another. This generalized the concept of "being more informative" which was introduced by Bohnenblust, Shapley, and Sherman [4] and may be found in Blackwell [1]. "Being more informative for k -decision problems" was introduced by Blackwell in [2]. ϵ -deficiency for k -decision problems was considered by the author in [19].

Let $\mathcal{G} = ((X, \mathcal{A}), (P_\theta; \theta \in \Theta))$ and $\mathcal{F} = ((Y, \mathcal{B}), (Q_\theta; \theta \in \Theta))$ be two experiments with the same parameter set Θ and let $\theta \rightsquigarrow \epsilon_\theta$ be a non-negative function on Θ (and let $k \geq 2$ be an integer).

Then we shall say that \mathcal{G} is ϵ -deficient relative to \mathcal{F} (for k -decision problems *) if to each decision space $^{**}(D, \mathcal{G})$

* When $k=2$: testing problems. ** i.e., a measurable space.

where \mathcal{J} is finite (where \mathcal{J} contains at most 2^k sets), every bounded loss-function *** $(\theta, d) \rightsquigarrow W_\theta(d)$ on $\Theta \times D$ and every risk function r obtainable in \mathcal{F} there is a risk function r' obtainable in \mathcal{G} so that

$$r'(\theta) \leq r(\theta) + \epsilon_\theta \|W\|, \quad \theta \in \Theta$$

where $\|W\| = \sup_{\theta, d} |W_\theta(d)|$.

If $(P_\theta; \theta \in \Theta)$ is dominated then ϵ -deficiency (for k -decision problems) for all finite subsets of Θ implies - by weak compactness - ϵ -deficiency (for k -decision problems).

If \mathcal{G} is 0-deficient relative to \mathcal{F} (for k -decision problems) then we shall say that \mathcal{G} is more informative than \mathcal{F} (for k -decision problems) and write this $\mathcal{G} \geq \mathcal{F}$ ($\mathcal{G} \geq_k \mathcal{F}$).

If $\mathcal{G} \geq \mathcal{F}$ ($\mathcal{G} \geq_k \mathcal{F}$) and $\mathcal{F} \geq \mathcal{G}$ ($\mathcal{F} \geq_k \mathcal{G}$) then we shall say that \mathcal{G} and \mathcal{F} are equivalent (for k -decision problems) and write this $\mathcal{G} \sim \mathcal{F}$ ($\mathcal{G} \sim_k \mathcal{F}$). By proposition 8 in [19] and by weak compactness $\mathcal{G} \sim_2 \mathcal{F} \Leftrightarrow \mathcal{G} \sim_3 \mathcal{F} \Leftrightarrow \dots \Leftrightarrow \mathcal{G} \sim \mathcal{F}$ provided \mathcal{G} and \mathcal{F} are dominated.

The greatest lower bound of all constants ϵ such that \mathcal{G} is ϵ -deficient relative to \mathcal{F} (for k -decision problems) will be denoted by $\delta(\mathcal{G}, \mathcal{F})$, respectively: $\delta_k(\mathcal{G}, \mathcal{F})$ and $\max[\delta(\mathcal{G}, \mathcal{F}), \delta(\mathcal{F}, \mathcal{G})]$ respectively: $\max[\delta_k(\mathcal{G}, \mathcal{F}), \delta_k(\mathcal{F}, \mathcal{G})]$ will be denoted by $\Delta(\mathcal{G}, \mathcal{F})$ respectively: $\Delta_k(\mathcal{G}, \mathcal{F})$.

If \mathcal{G} , \mathcal{F} and \mathcal{J} are experiments then:

*** It is always to be understood that $d \rightsquigarrow W_\theta(d)$ is measurable for each θ .

$$0 \leq \delta_k(\mathcal{E}, \mathcal{F}) \leq \delta_{k+1}(\mathcal{E}, \mathcal{F}) \leq \delta(\mathcal{E}, \mathcal{F}), \quad \delta(\mathcal{E}, \mathcal{E}) = 0, \delta_k(\mathcal{E}, \mathcal{F}) \uparrow \delta(\mathcal{E}, \mathcal{F}) \text{ as } k \rightarrow \infty, \text{ and}$$

$\delta_k(\mathcal{E}, \mathcal{F}) \leq \delta_k(\mathcal{E}, \mathcal{G}) + \delta_k(\mathcal{G}, \mathcal{F})$ so that $\Delta_2, \Delta_3, \dots, \Delta$ are all pseudometrics.

Let $\mathcal{E} = ((X, \mathcal{A}), (P_\theta; \theta \in \Theta))$ and $\mathcal{F} = ((Y, \mathcal{B}), (Q_\theta; \theta \in \Theta))$ be two experiments such that :

- (i) P_θ ; $\theta \in \Theta$ is dominated
- (ii) \mathcal{Y} is a Borel-sub set of a Polish space and \mathcal{B} is the class of Borel sub sets of \mathcal{Y} .

It follows from theorem 3 in Le Cam's paper [12] that \mathcal{E} is ϵ -deficient w.r.t. \mathcal{F} if and only if there is a Markov kernel M from (X, \mathcal{A}) to (Y, \mathcal{B}) so that $\|MP_\theta - Q_\theta\| \leq \epsilon_\theta$; $\theta \in \Theta$.

It was shown in [19] that :

$$\delta_k(\mathcal{E}, \mathcal{F}) = \sup_{\gamma \in \Gamma_k} [\gamma(\mathcal{F}) - \gamma(\mathcal{E})]$$

while

$$\delta(\mathcal{E}, \mathcal{F}) = \sup_{\gamma \in \Gamma} [\gamma(\mathcal{F}) - \gamma(\mathcal{E})]$$

when Θ is finite.

As indicated in Le Cam's paper [12] we may also consider weighted deficiencies.

Let λ be a prior distribution on Θ with countable support. For each ordered pair $(\mathcal{E}, \mathcal{F})$ of experiments we define the λ -weighted deficiency $\delta(\mathcal{E}, \mathcal{F} | \lambda)$ ($\delta_k(\mathcal{E}, \mathcal{F} | \lambda)$) as the greatest lower bound of all numbers $\sum_{\theta} \lambda_{\theta} \epsilon_{\theta}$ where \mathcal{E} is ϵ -deficient w.r.t. \mathcal{F} (for k -decision problems).

Trivially $|\delta(\mathcal{G}, \mathcal{F} | \lambda) - \delta(\mathcal{G}, \mathcal{F} | \mu)| \leq 2\|\mu - \lambda\|$ so that $\delta(\mathcal{G}, \mathcal{F} | \lambda)$ is uniformly continuous in λ . It follows directly from the definition that $\delta(\mathcal{G}, \mathcal{F} | \lambda)$ is concave in λ .

Suppose \mathcal{G} is dominated. By theorem 3 in Le Cam's paper [12] :

$$\delta(\mathcal{G}, \mathcal{F} | \lambda) = \inf_M \int \|P_\theta M - Q_\theta\| \lambda(d\theta)$$

where the inf is taken over all Markov operators from the band generated by $(P_\theta; \theta \in \Theta)$ to the band generated by $(Q_\theta; \theta \in \Theta)$.

Clearly $\delta_{(k)}(\mathcal{G}, \mathcal{F} | \lambda) \leq \delta_{(k)}(\mathcal{G}, \mathcal{F})$ for all λ and, by a simple minimax argument :

$$\delta_{(k)}(\mathcal{G}, \mathcal{F}) = \sup_\lambda \delta_{(k)}(\mathcal{G}, \mathcal{F} | \lambda)$$

when \mathcal{G} is dominated. Here the sup may be taken over all λ 's with finite support. If Θ is finite then the sup is obtained.

It is easily checked that :

$$\delta_{(k)}(\mathcal{G}, \mathcal{G} | \lambda) = 0 \quad \text{and that}$$

$$\delta_{(k)}(\mathcal{G}, \mathcal{H} | \lambda) \leq \delta_{(k)}(\mathcal{G}, \mathcal{F} | \lambda) + \delta_{(k)}(\mathcal{F}, \mathcal{H} | \lambda)$$

for experiments \mathcal{G} , \mathcal{F} and \mathcal{H} .

It follows that if we define $\Delta_{(k)}(\mathcal{G}, \mathcal{F} | \lambda) = \delta_{(k)}(\mathcal{G}, \mathcal{F} | \lambda) \vee \delta_{(k)}(\mathcal{F}, \mathcal{G} | \lambda)$ then $\Delta_2(\dots | \lambda), \dots, \Delta(\dots | \lambda)$ are all pseudo distances.

Assuming \mathcal{G} and \mathcal{F} are dominated, the distance Δ may be expressed as :

$$\Delta_{(k)}(\mathcal{G}, \mathcal{F}) = \sup_{\lambda} \Delta_{(k)}(\mathcal{G}, \mathcal{F} | \lambda) .$$

Lower bounds for $\delta(\mathcal{G}, \mathcal{F} | \lambda)$ may sometimes be found by noting that for any $\Theta_0 \subseteq \Theta$ such that $\lambda(\Theta_0) > 0$ we have

$$\delta_{(k)}(\mathcal{G}, \mathcal{F} | \lambda) \geq \lambda(\Theta_0) \delta_{(k)}(\mathcal{G}_{\Theta_0}, \mathcal{F}_{\Theta_0} | \lambda_{\Theta_0})$$

where λ_{Θ_0} is the conditional distribution given Θ_0 when θ is distributed according to λ .

The following facts may reduce the problem of evaluating a deficiency. Suppose $\delta(\mathcal{G}, \mathcal{F}) = \delta(\mathcal{G}, \mathcal{F} | \mu)$. By the result of Le Cam cited above there is a Markov operator such that $\delta(\mathcal{G}, \mathcal{F}) = \sup_{\theta} \|P_{\theta} M - Q_{\theta}\|$. Then $\|P_{\theta} M - Q_{\theta}\| = \delta(\mathcal{G}, \mathcal{F})$ whenever $\mu_{\theta} > 0$.

Suppose on the other hand that

$$\sum_{\theta} \mu_{\theta} \|P_{\theta} M - Q_{\theta}\| = \delta(\mathcal{G}, \mathcal{F} | \mu) \quad \text{and that}$$

$$\begin{aligned} \|P_{\theta} M - Q_{\theta}\| &= \sup_{\theta} \|P_{\theta} M - Q_{\theta}\| \quad \text{when } \mu_{\theta} > 0 . \quad \text{Then } \delta(\mathcal{G}, \mathcal{F}) = \delta(\mathcal{G}, \mathcal{F} | \mu) \\ &= \sup_{\theta} \|P_{\theta} M - Q_{\theta}\| . \end{aligned}$$

If \mathcal{G} and \mathcal{F} exhibit symmetry properties then, under certain conditions, we may restrict attention to Markov operators which exhibit the same symmetries. We refer the reader to Boll [5], Le Cam [12], Heyer [10] and the author [20] for further information on this. It suffices for the purposes of this paper to note that if G is a finite group of transformations on Θ such

that $(P_{g(\theta)}; \theta \in \Theta) \sim (P_\theta; \theta \in \Theta)$ and $(Q_{g(\theta)}; \theta \in \Theta) \sim (Q_\theta; \theta \in \Theta)$ when

$$g \in G \text{ then, by concavity, } \delta_{(k)}(\mathcal{G}, \mathcal{F}) = \sup_{\lambda} \delta_{(k)}(\mathcal{G}, \mathcal{F} | \lambda)$$

where the sup is taken over all G invariant prior distributions with finite support. In particular, if Θ is finite and \mathcal{G} and \mathcal{F} are both completely symmetric then

$$\delta_{(k)}(\mathcal{G}, \mathcal{F}) = \delta_{(k)}(\mathcal{G}, \mathcal{F} | \mu)$$

where μ is the uniform distribution on Θ . We may also note that $\delta_{(k)}(\mathcal{G}, \mathcal{F}) = \delta_{(k)}(\mathcal{G}_{\Theta_0}, \mathcal{F}_{\Theta_0})$ if $\delta_{(k)}(\mathcal{G}, \mathcal{F}) = \delta_{(k)}(\mathcal{G}, \mathcal{F} | \mu)$ and $\mu(\theta) = 0$ when $\theta \notin \Theta_0$.

The notations \wedge and \vee are used on several occasions for inf and sup. If, in particular, μ_t ; $t \in T$ are measures then $\bigwedge_t \mu_t$ and $\bigvee_t \mu_t$ are, respectively, notations for $\inf_t \mu_t$ and $\sup_t \mu_t$ for the family $\{\mu_t$; $t \in T\}$ w.r.t. the set wise ordering of measures.

If μ is a measure then $\|\mu\|$ denotes the total variation of μ ; i.e. $\|\mu\| = \sup \left\{ \int f d\mu : -1 \leq f \leq 1 \right\}$.

We will reserve the letter Λ as a notation for the set of all prior distributions with finite support. The letter \mathcal{L} will be used as a notation for "distribution of". Thus, for example, $\mathcal{L}_P(f)$ is the distribution of f under P .

3. Deficiencies w.r.t. a totally informative experiment.

Consider the problem of estimating the unknown parameter θ when the loss is 0 or 1 as the estimator hits or fails. The estimator should be based on an experiment $\mathcal{G} = (P_\theta; \theta \in \Theta)$ and we will assume, for reasons to be explained later, that Θ is countable. We shall admit randomized estimators. Thus an estimator δ is a rule which to each point x in the sample space of \mathcal{G} assigns a probability measure $\delta_\theta(x); \theta \in \Theta$ on Θ such that, for fixed θ , $\delta_\theta(x)$ is measurable in x . The operating characteristic of δ is the rule which to each $\theta \in \Theta$ assigns the probability distribution $E_\theta \delta_\theta; \theta' \in \Theta$ on Θ . The risk, $r_\delta(\theta)$, is given by:

$$r_\delta(\theta) = 1 - E_\theta \delta_\theta; \theta \in \Theta.$$

If λ is any prior distribution, then the Baye's risk of δ is $1 - \sum_\theta \lambda_\theta E_\theta \delta$. The minimum Baye's risk for the prior λ will be denoted as $b(\lambda | \mathcal{G})$. Simple calculations show that:

$$b(\lambda | \mathcal{G}) = 1 - \|\bigwedge_\theta \lambda_\theta P_\theta\|$$

If $\Theta = \{\theta_1, \theta_2\}$ then this may ^{*)} be rewritten as:

$$b(\lambda | \mathcal{G}) = \|\bigwedge_\theta \lambda_\theta P_\theta\|.$$

*) We may note in passing that, for any countable Θ , $\|\bigwedge_\theta \lambda_\theta P_\theta\|$ is the minimum Baye's risk for the prior λ w.r.t. the decision problem (T, L) where T consists of all sub sets of Θ which are complements of one point sets and $L_\theta(t) = 0$ or 1 as $\theta \in t$ or $\theta \notin t$.

If each P_θ has a density f_θ w.r.t. some measure σ , then an estimator δ achieves minimum Baye's risk if and only if, for σ almost all x , $\delta(x)$ is supported by $\{\theta : \lambda_\theta f_\theta(x) = \max_{\theta' \in \Theta} \lambda_{\theta'} f_{\theta'}(x)\}$. If Θ is finite then we find, in particular, that any maximum likelihood estimator of θ achieves minimum Baye's risk for the uniform distribution.

Comparison of an experiment w.r.t. the totally informative experiment may be completely described in terms of this decision problem:

Theorem 3.1.

For any experiment \mathcal{E} :

$$\delta(\mathcal{E}, \mathcal{M}_a | \lambda) = 2b(\lambda | \mathcal{E}) .$$

Corollary 3.2.

For any experiment \mathcal{E} :

$$\delta(\mathcal{E}, \mathcal{M}_a) = 2 \sup_{\lambda} b(\lambda | \mathcal{E}) .$$

Remark:

By theorem 3.1 and Morse's and Sacksteder's paper [18] the function $\lambda \rightsquigarrow b(\mathcal{E}, \mathcal{M}_a | \lambda)$ defines the experiment \mathcal{E} up to equivalence. The proof of theorem 3.1 follows immediately from:

Theorem 3.3.

\mathcal{E} is ϵ -deficient w.r.t. \mathcal{M}_a if and only if there is a decision rule δ so that

$$1 - E_\theta \delta_\theta \leq \epsilon_\theta / 2 ; \theta \in \Theta .$$

Proof:

\mathcal{G} is ϵ -deficient w.r.t. $\mathcal{M}_a \Leftrightarrow$ there exists an estimator M such that $\|P_\theta M - \text{one point distribution in } \theta\| \leq \epsilon_\theta$; $\theta \in \Theta \Leftrightarrow$ there exists an estimator M so that $1 - E_\theta M_\theta \leq \epsilon_\theta/2$.

□

Corollary 3.4.

Suppose P_θ ; $\theta \in \Theta$ are non atomic and that Θ is finite. Then $(P_\theta; \theta \in \Theta)$ is ϵ -deficient w.r.t. \mathcal{M}_a if and only if there are disjoint sets A_θ ; $\theta \in \Theta$ in the sample space so that $P_\theta(A_\theta) \equiv 1 - \frac{\epsilon_\theta}{2}$.

Proof:

This follows from Dvoretzky's, Wald's and Wolfowitz's extension [8] of Ljapounov's theorem [17] .

□

By the last theorem $\delta(\mathcal{G}, \mathcal{M}_a)/2$ may be characterized by two properties :

$$(i) \quad \text{For some estimator } \delta \\ \sup_{\theta} P_{\theta, \delta}(\delta \neq \theta) = \frac{\delta(\mathcal{G}, \mathcal{M}_a)}{2}$$

$$(ii) \quad \text{For all estimators } \delta \\ \sup_{\theta} P_{\theta, \delta}(\delta \neq \theta) \geq \frac{\delta(\mathcal{G}, \mathcal{M}_a)}{2}$$

In other words : $\delta(\mathcal{G}, \mathcal{M}_a)/2$ is the minimax risk for the problem of guessing θ with 0-1 loss.

Call a prior distribution μ least favorable when $\delta(\mathcal{C}, \mathcal{U}_a | \mu) = \delta(\mathcal{C}, \mathcal{U}_a)$. Thus a prior distribution is least favorable if and only if it is least favorable for the decision problem.

The following facts follow then readily from general principles of statistical decision theory :

Proposition 3.5.

(i) Let μ be least favorable and let δ achieve the minimax risk. Then $P_{\theta, \delta}(\delta \neq \theta) = \delta(\mathcal{C}, \mathcal{U}_a)/2$ when $\mu(\theta) > 0$.

(ii) Suppose δ achieves minimum Baye's risk for the prior μ and that $P_{\theta, \delta}(\delta \neq \theta) = \max_{\theta'} P_{\theta', \delta}(\delta \neq \theta')$ when $\mu(\theta) > 0$. Then μ is least favorable and δ is a minimax procedure.

The quantities introduced above may also be interpreted in terms of probabilities of guessing the right value of θ . Thus $1 - b(\lambda | \mathcal{C}) = 1 - \frac{1}{2}\delta(\mathcal{C}, \mathcal{U}_a | \lambda)$ equals the maximum probability of guessing θ for the prior λ . Similarly $1 - \frac{1}{2}\delta(\mathcal{C}, \mathcal{U}_a)$ is the maximum of the minimum probability of guessing θ for all guessing procedures.

Let us consider a few simple inequalities for these quantities :

Proposition 3.6.

Let λ be a prior distribution and Θ_0 a sub set of Θ so that $\lambda(\Theta_0) > 0$.

Then :

$$b(\lambda | \mathcal{C}) \geq \lambda(\Theta_0) b(\lambda_{\theta} / \lambda(\Theta_0) ; \theta \in \Theta_0 | \mathcal{C}_{\Theta_0})$$

or equivalently :

$$\delta(\mathcal{E}, \mathcal{M}_a | \lambda) \geq \lambda(\Theta_0) \delta(\mathcal{E}_{\Theta_0}, \mathcal{M}_a | \lambda_{\Theta_0} / \lambda(\Theta_0)) ; \theta \in \Theta_0$$

Proof :

This follows directly from theorem 3.1 and section 2.

□

Proposition 3.7.

For any prior distribution λ :

$$2b(\lambda | \mathcal{E}) \leq \delta(\mathcal{E}, \mathcal{M}_a) \leq 2[\#\Theta]b \text{ (uniform distribution } |\Theta) \text{ .}$$

Comparison may occasionally be reduced to pairs by

Proposition 3.8.

Let λ be a non degenerate prior distribution.

Then :

$$2b(\lambda | \mathcal{E}) \leq \sum_{\theta_1 \neq \theta_2} (\lambda_{\theta_1} + \lambda_{\theta_2}) b\left(\left(\frac{\lambda_{\theta_1}}{\lambda_{\theta_1} + \lambda_{\theta_2}}, \frac{\lambda_{\theta_2}}{\lambda_{\theta_1} + \lambda_{\theta_2}}\right) | \mathcal{E}_{\{\theta_1, \theta_2\}}\right)$$

Hence :

$$\delta(\mathcal{E}, \mathcal{M}_a) \leq \sum_{\theta_1 \neq \theta_2} \delta(\mathcal{E}_{\{\theta_1, \theta_2\}}, \mathcal{M}_a)$$

Corollary 3.9.

Suppose $\#\Theta = m < \infty$. Then :

$$\delta(\mathcal{E}, \mathcal{M}_a) \leq (m-1) \max_{\theta_1 \neq \theta_2} \delta(\mathcal{E}_{\{\theta_1, \theta_2\}}, \mathcal{M}_a)$$

and

$$\begin{aligned} b(\text{uniform } |\mathcal{G}) &\leq \frac{1}{m} \sum_{\theta_1 \neq \theta_2} b((\tfrac{1}{2}, \tfrac{1}{2}) | \mathcal{G}_{\{\theta_1, \theta_2\}}) \\ &\leq (m-1) \max_{\theta_1 \neq \theta_2} b((\tfrac{1}{2}, \tfrac{1}{2}) | \mathcal{G}_{\{\theta_1, \theta_2\}}) . \end{aligned}$$

Proof of proposition 3.8 :

Let a_n ; $n=1,2,\dots$ be a bounded sequence of positive numbers.
Then

$$\sum_n a_n - \vee_n a_n \leq \frac{1}{2} \sum_{m \neq n} a_m \wedge a_n$$

Hence:

$$b(\lambda | \mathcal{G}) = \|\sum_{\theta} \lambda_{\theta} P_{\theta} - \vee_{\theta} \lambda_{\theta} P_{\theta}\| \leq \frac{1}{2} \sum_{\theta_1 \neq \theta_2} \|\lambda_{\theta_1} P_{\theta_1} \wedge \lambda_{\theta_2} P_{\theta_2}\| .$$

The proof follows now from the formula: $b((\lambda_1, \lambda_2) | (P_1, P_2)) = \|\lambda_1 P_1 \wedge \lambda_2 P_2\|$.

□

If Θ is finite then, by [19] :

$$\delta(\mathcal{G}, \mathcal{M}_a) = \sup_{\gamma \in \Gamma} [1 - \int \gamma(dP_{\theta}; \theta \in \Theta)]$$

where Γ is the class of all sub linear functions on R^{Θ} such that $\gamma(-e^{\theta}) \equiv \gamma(e^{\theta})$ and $\sum_{\theta} \gamma(e^{\theta}) = 1$.

It follows from theorem 3.1 that it suffices to consider functions γ of the form : $\gamma(x) = 2 \vee_{\theta} \lambda_{\theta} x_{\theta} - \sum_{\theta} \lambda_{\theta} x_{\theta}$.

Restricting ρ to the sub class Γ_k of Γ consisting of those functions in Γ which are pointwise maxima of k linear functionals we get, [19], the formula :

$$\delta_k(\mathcal{E}, \mathcal{M}_a) = \sup_{\gamma \in \Gamma_k} [1 - \int \gamma(dP_\theta : \theta \in \Theta)]$$

The case of comparison by testing problems is of particular interest. It follows from section 3 in [19] that :

Proposition 3.10.

\mathcal{E} is ϵ -deficient w.r.t. \mathcal{M}_a for testing problems if and only if there to each subset Θ_0 of Θ corresponds a power-function π in \mathcal{E} so that $\pi(\theta) \leq \epsilon_\theta/2$ or $\geq 1 - \epsilon_\theta/2$ as $\theta \in \Theta_0$ or $\theta \notin \Theta_0$.

The sublinear function criterion yields :

Proposition 3.11.

$$\delta_2(\mathcal{E}, \mathcal{M}_a) = \sup_{\theta} [1 - \|\sum a_\theta P_\theta\|]$$

where the sup taken over all real valued functions a on Θ having finite support and satisfying : $\sum_\theta |a_\theta| = 1$.

Let Θ be infinite and not countable and suppose $\mathcal{E} = (P_\theta; \theta \in \Theta)$ is dominated. Put $\mathcal{M}_a = (Q_\theta; \theta \in \Theta)$ where, for each θ , Q_θ is the one point distribution in θ . By the randomization criterion there is a dominated family $(\tilde{P}_\theta; \theta \in \Theta)$ of probability measures such that $\|\tilde{P}_\theta - Q_\theta\| \leq \delta(\mathcal{E}, \mathcal{M}_a)$; $\theta \in \Theta$. Thus $\tilde{P}_\theta(\theta) \geq 1 - \delta(\mathcal{E}, \mathcal{M}_a)/2$ for all θ and this implies, since \mathcal{E} is dominated, that $\delta(\mathcal{E}, \mathcal{M}_a) = 2$.

In order to get interesting results in the case of non countable parameter sets, we shall have to restrict ourselves to non dominated experiments. This is the main reason for excluding non countable parameter sets in this investigation.

Example 3.12.

Let for each $\xi \in [0, 1]^\Theta$, $\mathcal{E}_\xi = (P_\theta; \theta \in \Theta)$, where P_θ assigns probabilities $1 - \xi_\theta$ and ξ_θ to, respectively, θ and Θ . The experiments \mathcal{E}_ξ have many interesting properties and we shall here consider a few of them.

It was shown in [21] that they belong to the larger (but not "much" larger) family of experiments \mathcal{E} such that $\mathcal{F} \geq \mathcal{E}$ if and only if $\mathcal{F} \sim \mathcal{E} \times \mathcal{G}$ for some experiment \mathcal{G} .

They are closed under products since $\mathcal{E}_\xi \times \mathcal{E}_\eta = \mathcal{E}_{\xi\eta}$. This follows readily from the fact that $H(t|\mathcal{E}_\xi) = \prod_{\theta} \xi_\theta^{t_\theta}$ for each prior distribution with finite support.

The last property implies that $\mathcal{E}_\xi \geq \mathcal{E}_\eta$ whenever $\xi \leq \eta$. Suppose, conversely, that $\mathcal{E}_\xi \geq \mathcal{E}_\eta$. By the sublinear function criterion, [19], we find that

$$(\xi_{\theta_0} - \beta \xi_{\theta_1})^+ - \xi_{\theta_0} \geq (\eta_{\theta_0} - \beta \eta_{\theta_1})^+ - \eta_{\theta_0}$$

for any pair (θ_0, θ_1) of **distinct** points and any $\beta > 0$. $\beta \rightarrow \infty$, yield, provided $\xi_{\theta_1} > 0$, that $\xi_{\theta_0} \leq \eta_{\theta_0}$. Thus $\xi \leq \eta$ provided

$$\mathcal{G}_\xi \vdash \mathcal{M}_a. \text{ Note that we only employed functions } \gamma \in \Gamma_2.$$

It follows easily, that :

$$\mathcal{G}_\xi \geq \mathcal{G}_\eta \Leftrightarrow \mathcal{G}_\xi \geq \frac{\mathcal{G}_\eta}{2} \Leftrightarrow H(t|\mathcal{G}_\xi) \leq H(t|\mathcal{G}_\eta)$$

for all prior distributions t with finite support.

This conditions are in turn, provided $\mathcal{G}_\xi \vdash \mathcal{M}_a$ equivalent with the condition $\xi \leq \eta$. [It is easily seen that $\mathcal{G}_\xi \sim \mathcal{M}_a$ if and only if $\xi_\theta > 0$ for at most one θ .]

Let us compute a few deficiencies for these experiments. If Θ has a finite number m of elements then, by the sublinear function criterion :

$$\begin{aligned} \delta_{(k)}(\mathcal{G}_\xi, \mathcal{G}_\eta) &= \sup_{\gamma \in \Gamma_{(k)}} [\gamma(\mathcal{G}_\eta) - \gamma(\mathcal{G}_\xi)] \\ &= \sup_{\gamma \in \Gamma_{(k)}} [\gamma(\eta) - \gamma(\xi) - \sum_{\theta} (\eta_\theta - \xi_\theta) \gamma(e^\theta)] . \end{aligned}$$

In particular, using the expressions for $\delta_k(\mathcal{M}_i, \mathcal{M}_a)$ in [19] :

$$\begin{aligned} \delta_k(\mathcal{G}_{\alpha, \alpha, \dots, \alpha}, \mathcal{G}_{\beta, \beta, \dots, \beta}) &= \sup_{\gamma \in \Gamma_k} (\beta - \alpha) [\gamma(e) - \sum_{\theta} \gamma(e^\theta)] \\ &= (\alpha - \beta)^+ \delta_k(\mathcal{M}_i, \mathcal{M}_a) = 2(\alpha - \beta)^+ (1 - \frac{1}{m \wedge k}) \end{aligned}$$

$k \rightarrow \infty$ yield :

$$\delta(\mathcal{E}_{\alpha, \dots, \alpha}, \mathcal{E}_{\beta, \dots, \beta}) = 2(\alpha - \beta)^+(1 - \frac{1}{m}) .$$

Letting $m \rightarrow \infty$ we see that in any case :

$$\delta_k(\mathcal{E}_{\alpha, \alpha, \dots}, \mathcal{E}_{\beta, \beta, \dots}) = 2(\alpha - \beta)^+(1 - \frac{1}{[\# \Theta] \wedge k})$$

and

$$\delta(\mathcal{E}_{\alpha, \alpha, \dots}, \mathcal{E}_{\beta, \beta, \dots}) = 2(\alpha - \beta)^+(1 - \frac{1}{[\# \Theta]})$$

In particular :

$$\delta_k(\mathcal{E}_{\alpha, \alpha, \dots}, \mathcal{M}_a) = 2\alpha(1 - \frac{1}{[\# \Theta] \wedge k})$$

$$\delta(\mathcal{E}_{\alpha, \alpha, \dots}, \mathcal{M}_a) = 2\alpha(1 - \frac{1}{[\# \Theta]})$$

$$\delta_k(\mathcal{M}_i, \mathcal{E}_{\beta, \beta, \dots}) = 2(1 - \beta)(1 - \frac{1}{[\# \Theta] \wedge k})$$

and

$$\delta(\mathcal{M}_i, \mathcal{E}_{\beta, \beta, \dots}) = 2(1 - \beta)(1 - \frac{1}{[\# \Theta]}) .$$

We shall need an expression for $\delta(\mathcal{E}_{\xi_1, \xi_2, \xi_3, \dots}, \mathcal{M}_a)$ for a general ξ . Put for each r -tuple (z_1, z_2, \dots, z_r) of non negative numbers :

$$H(z_1, z_2, \dots, z_r) = [r^{-1} \sum_{i=1}^r z_i^{-1}] \text{ or } = 0$$

as $z_1 \cdot z_2 \dots \cdot z_r > 0$ or $z_1 \cdot z_2 \dots \cdot z_r = 0$.

Thus $H(z_1, z_2, \dots, z_r)$ is just the harmonic mean of the numbers (z_1, z_2, \dots, z_r) . Then :

$$\delta(\mathcal{E}_\xi, \mathcal{M}_a | \lambda) / 2 = b(\lambda | \mathcal{E}_\xi) = \sum_{\theta} \lambda_{\theta} \xi_{\theta} - \sqrt{\lambda_{\theta} \xi_{\theta}}$$

and

$$\delta(\mathcal{E}_\xi, \mathcal{M}_a) / 2 = \sup_{r: r \leq \# \Theta} (1 - \frac{1}{r}) \sup H(\xi_{\theta_1}, \xi_{\theta_2}, \dots, \xi_{\theta_r})$$

where the last sup is taken over all r point sub sets $\{\theta_1, \theta_2, \dots, \theta_r\}$ of Θ .

Proof :

Put $\mathcal{E} = (P_{\theta}; \theta \in \Theta)$. Then $b(\lambda) \stackrel{\text{def.}}{=} b(\lambda | \mathcal{E}_\xi) = 1 - \|\sqrt{\lambda_{\theta}} P_{\theta}\|$

$$= 1 - [\sum_{\theta} \lambda_{\theta} (1 - \xi_{\theta}) + \sum_{\theta} \lambda_{\theta} \xi_{\theta}]$$

$$= \sum_{\theta} \lambda_{\theta} \xi_{\theta} - \sqrt{\lambda_{\theta} \xi_{\theta}}.$$

The formula for $\delta(\mathcal{E}_\xi, \mathcal{M}_a | \lambda)$ follows now from theorem 3.1. By corollary 3.2 : $\delta(\mathcal{E}_\xi, \mathcal{M}_a) / 2 = \sup_{\lambda} b(\lambda)$. The expression for $\delta(\mathcal{E}_\xi, \mathcal{M}_a)$ when Θ is infinite follows directly from the expression for $\delta(\mathcal{E}_\xi, \mathcal{M}_a)$ when Θ is finite. We may therefore, without loss of generality, assume that Θ is finite. Note then that $b(\lambda)$ is simply the sum of all numbers $\lambda_{\theta} \xi_{\theta}$ except one of the largest. How do we maximize this sum ? Let, for each non empty sub set U of Θ , $C(U)$ be the set of prior distributions λ such that $\lambda_{\theta} \xi_{\theta} = \max_{\theta' \in U} \lambda_{\theta'} \xi_{\theta'}$, when $\theta \in U$. Then $C(U)$ is compact, convex and b is affine on $C(U)$. Furthermore any prior distribution belongs to some set $C(U)$. Hence :

$$\max_{\lambda} b(\lambda) = \max_U \max_{\lambda \in C(U)} b(\lambda) \quad \text{and} \quad \max \{b(\lambda) : \lambda \in C(U)\} =$$

$\max \{b(\lambda) : \lambda \in \text{ext } C(U)\}$ where $\text{ext } C(U)$ is the set of extreme points of $C(U)$.

Consider a point $\lambda^0 \in C(U)$. Put $W = \{\theta : \lambda_{\theta}^0 \xi_{\theta} = \vee_{\theta} \lambda_{\theta}^0 \xi_{\theta}\}$ and

let $\theta_1 \notin W$. Suppose $M = \max_{\theta'} \lambda_{\theta'}^0 \xi_{\theta'} > 0$ and that $\lambda_{\theta_1}^0 > 0$. Put :

$$\lambda_{\theta}^h = \lambda_{\theta}^0 - h \xi_{\theta}^{-1} [\sum_W \xi_{\theta}^{-1}]^{-1} ; \quad \theta \in W$$

$$\lambda_{\theta_1}^h = \lambda_{\theta_1}^0 + h$$

$$\lambda_{\theta}^h = \lambda_{\theta}^0 ; \quad \theta \notin W \cup \{\theta_1\} .$$

Then, provided h is sufficiently small in absolute value, $\lambda^h \in C(U)$. Hence, since $\lambda^0 = \frac{1}{2}(\lambda^h + \lambda^{-h})$, $\lambda^0 \notin \text{ext } C(U)$.

It follows that $\lambda_{\theta} \xi_{\theta} = \max_{\theta'} \lambda_{\theta'} \xi_{\theta'}$, for all θ such that $\lambda_{\theta} > 0$, when $\lambda \in \text{ext } C(U)$.

If $\max_{\theta'} \lambda_{\theta'} \xi_{\theta'} > 0$, $r = \#\{\theta : \lambda_{\theta} \xi_{\theta} = \vee_{\theta} \lambda_{\theta} \xi_{\theta}\}$ and $\lambda \in \text{ext } C(U)$

then $b(\lambda) = (1 - \frac{1}{r}) H(\xi_{\theta_1}, \xi_{\theta_2}, \dots, \xi_{\theta_r})$ where $\{\theta_1, \theta_2, \dots, \theta_r\} =$

$\{\theta : \lambda_{\theta} \xi_{\theta} = \vee_{\theta} \lambda_{\theta} \xi_{\theta}\}$.

If $\max_{\theta'} \lambda_{\theta'} \xi_{\theta'} = 0$, then $b(\lambda) = 0 = (1 - \frac{1}{r}) H(\xi_{\theta_1}, \dots, \xi_{\theta_r})$ for some sub set $\{\theta_1, \dots, \theta_r\}$ of Θ . It follows that

$$\delta(\mathcal{C}_5, \mathcal{U}_a)/2 \leq \sup_r (1 - \frac{1}{r}) H(\xi_{\theta_1}, \dots, \xi_{\theta_r}) .$$

Conversely, any positive number $(1 - \frac{1}{r}) H(\xi_{\theta_1}, \dots, \xi_{\theta_r}) = b(\lambda)$ where $\lambda_{\theta_i} = \xi_{\theta_i}^{-1} [\sum_{i=1}^r \xi_{\theta_i}^{-1}]^{-1}$. Thus $(1 - \frac{1}{r}) H(\xi_{\theta_1}, \dots, \xi_{\theta_r}) \geq \delta(\mathcal{E}_\xi, \mathcal{M}_a)/2$ for any r -point sub set $\{\theta_1, \dots, \theta_r\}$ of Θ . Altogether this proves the formula for $\delta(\mathcal{E}_\xi, \mathcal{M}_a)/2$.

It follows readily that for any countable parameter set :

$$H(a, b) \leq \delta(\mathcal{E}_\xi, \mathcal{M}_a) \leq 2 H(a, b)$$

where $a = \sup_{\theta} \xi_{\theta}$ and $b = \sup_{\theta_1 \neq \theta_2} \xi_{\theta_1} \wedge \xi_{\theta_2}$.

If $\Theta = \{1, 2, \dots, m\}$ and $\xi_1 \geq \xi_2 \geq \dots \geq \xi_m$ then

$$\delta(\mathcal{E}_\xi, \mathcal{M}_a)/2 = \max_{r \leq m} (1 - \frac{1}{r}) H(\xi_1, \xi_2, \dots, \xi_r).$$

The distribution which assigns mass $\xi_i^{-1} \sum_{i=1}^r \xi_i^{-1}$ to

i , $i=1, 2, \dots, r$ is least favorable provided

$$\delta(\mathcal{E}_\xi, \mathcal{M}_a)/2 = (1 - \frac{1}{r}) H(\xi_1, \xi_2, \dots, \xi_r) > 0.$$

4. Comparison of experiments w.r.t. a totally non informative experiment.

Consider decision problems where the set T of possible decisions consists of all probability distributions on the sample space of $\mathcal{E} = (P_\theta; \theta \in \Theta)$ and the loss function L is given by :

$$L_\theta(M) = \|M - P_\theta\| ; \theta \in \Theta .$$

If no observations are available then the minimax risk in this estimation problem is $\inf_M \sup_{\theta \in \Theta} \|P_\theta - M\|$ and this is precisely the deficiency of \mathcal{M}_i w.r.t. \mathcal{E} . More generally :

Theorem 4.1.

\mathcal{M}_i is ϵ -deficient w.r.t. $\mathcal{E} = (P_\theta; \theta \in \Theta)$ if and only if there is a probability distribution Q so that

$$\|P_\theta - Q\| \leq \epsilon_\theta ; \theta \in \Theta .$$

Proof :

This follows immediately from the randomization criterion.

□

Corollary 4.2.

$$\frac{1}{2} \sup_{\theta_1 \neq \theta_2} \|P_{\theta_1} - P_{\theta_2}\| \leq \delta(\mathcal{M}_i, \mathcal{E}) \leq (1 - \frac{1}{\#\Theta}) \sup_{\theta_1 \neq \theta_2} \|P_{\theta_1} - P_{\theta_2}\|$$

If Θ is finite and \bar{P} is the average of the measures P_θ ; $\theta \in \Theta$ then $\max_{\theta} \|P_\theta - \bar{P}\|$ lies between $\delta(\mathcal{M}_i, \mathcal{E})$ and the extreme right of the above inequality. Furthermore we know, by theorem 4.3, that $\delta(\mathcal{M}_i, \mathcal{E})$ is exactly equal to this quantity when $\#\Theta = 2$. It might therefore be tempting to use this quantity as an upper bound. Unfortunately this bound may, as we now shall see, be very inaccurate.

Example 4.3.

Suppose $\Theta = \{1, 2, 3\}$ and that $\mathcal{E} = \mathcal{E}_{0, \epsilon, \epsilon}$ i.e. \mathcal{E} is given by the matrix :

$\theta \backslash x$	1	2	3	4
1	1	0	0	0
2	0	$1-\epsilon$	0	ϵ
3	0	0	$1-\epsilon$	ϵ

where $\epsilon \in [0, 1]$.

As $\#\Theta = 3$ any deficiency is in $[0, \frac{4}{3}]$. Hence, since $\|P_1 - \bar{P}\| = \frac{4}{3}$:

$$\max_{\theta} \|P_\theta - \bar{P}\| = \frac{4}{3} \quad \text{for all } \epsilon.$$

Put, for each distribution $Q = (Q(1), Q(2), Q(3), Q(4))$ on $\{1, 2, 3, 4\}$, $F(Q) = \max_{\theta} \|P_\theta - Q\|$. Hence $\delta(\mathcal{M}_i, \mathcal{E}) = \min_Q F(Q)$.

Simple calculations yield :

$$\|P_1 - Q\| = 2(1 - Q(1))$$

$$\|P_2 - Q\| = Q(1) + |1 - \epsilon - Q(2)| + Q(3) + |\epsilon - Q(4)|$$

and

$$\|P_3 - Q\| = Q(1) + Q(2) + |1 - \epsilon - Q(3)| + |Q(4) - \epsilon|.$$

It follows that F is symmetric in $Q(2)$ and $Q(3)$. Hence we may, by convexity assume that $Q(2) = Q(3)$. We may then write

$$F(Q) = \max[2(1 - Q(1)), B(Q)] \quad \text{where}$$

$$B(Q) = Q(1) + Q(2) + |1 - \epsilon - Q(2)| + |Q(4) - \epsilon|$$

$$= \frac{1}{2}[1 + Q(1) - Q(4) + |1 - 2\epsilon + Q(1) + Q(4)|] + |Q(4) - \epsilon|$$

where $Q(1) + Q(4) \leq 1$. We have utilized the equality :

$$Q(1) + 2Q(2) + Q(4) = 1. \quad \text{Distinguishing the cases } "Q(4) \geq \epsilon"$$

and $"Q(4) \leq \epsilon"$ we find by a simple convexity analysis that

$$\min_Q F(Q) = \frac{4}{3} - \frac{2\epsilon}{3} \quad \text{or} = 1 \quad \text{as } \epsilon \leq \frac{1}{2} \quad \text{or } \epsilon \geq \frac{1}{2}. \quad \text{In the first}$$

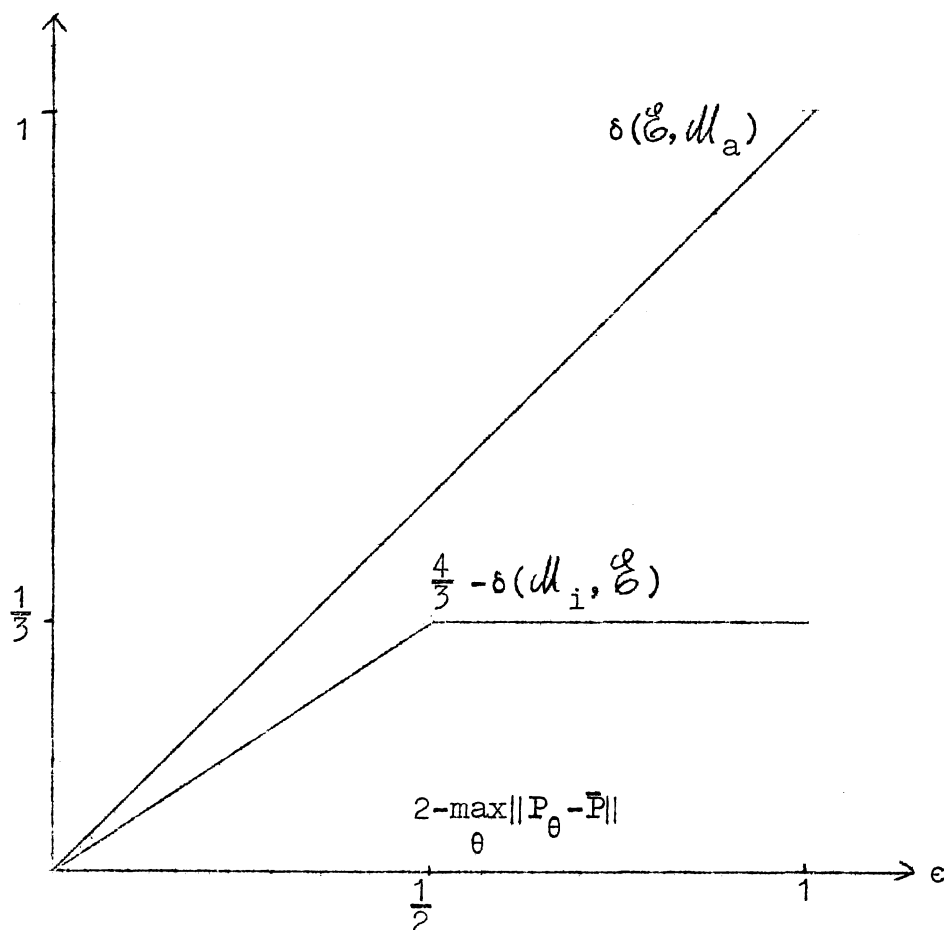
case the minimum is obtained for $Q(1) = \frac{1}{3}(\epsilon + 1)$, $Q(2) = Q(3) =$

$$= \frac{1}{3}(1 - 2\epsilon) \quad \text{and} \quad Q(4) = \epsilon. \quad \text{By section 3, } \delta(\mathcal{E}, \mathcal{M}_a) = \epsilon \quad \text{for}$$

all ϵ .

Let us compare the quantities, $\frac{4}{3} - \max_{\theta} \|P_{\theta} - \bar{P}\|$, $\frac{4}{3} - \delta(\mathcal{M}_i, \mathcal{E})$

and $\delta(\mathcal{E}, \mathcal{M}_a) = \epsilon$ as functions of ϵ .



We see immediately that there is no general inequality of the form : $H \delta(\mathcal{C}, \mathcal{M}_a) \leq \frac{4}{3} - \max_{\theta} \|P_{\theta} - \bar{P}\|$ where H is a positive constant. The figure indicates, however, the possibility of finding a positive constant H (here $H = \frac{2}{3}$ will do) such that $H \delta(\mathcal{C}, \mathcal{M}_a) \leq \frac{4}{3} - \delta(\mathcal{M}_i, \mathcal{C})$.

The left hand side of the inequality of Corollary 4.2 is a consequence of :

Theorem 4.4.

$$\delta_2(\mathcal{M}_i, \mathcal{G}) = \frac{1}{2} \sup_{\theta_1, \theta_2} \|P_{\theta_1} - P_{\theta_2}\| .$$

Thus $2 \delta_2(\mathcal{M}_i, \mathcal{G})$ is simply the diameter of \mathcal{G} for statistical distance.

Proof :

It suffices to consider the case of a finite Θ . By corollary 6 in [19] $\delta_2(\mathcal{M}_i, \mathcal{G}) \leq \epsilon$ if and only if to each measurable set A in the sample space of \mathcal{G} corresponds a number $t \in [0,1]$ so that $\inf_t \sup_{\theta} 2|P_{\theta}(A) - t| \leq \epsilon$. This is possible if and only if $|P_{\theta_1}(A) - P_{\theta_2}(A)| \leq \epsilon$ when $\theta_1, \theta_2 \in \Theta$. □

If $\#\Theta = 2$ then the deficiencies are given in :

Theorem 4.5.

Suppose $\Theta = \{1,2\}$ and $\mathcal{G} = (P_1, P_2)$.

Then :

$$\delta(\mathcal{M}_i, \mathcal{G}) = \delta(\mathcal{M}_i, \mathcal{G} | \frac{1}{2}, \frac{1}{2}) = \frac{1}{2} \|P_1 - P_2\| = 1 - 2b(\frac{1}{2}, \frac{1}{2} | \mathcal{G})$$

and

$$\delta(\mathcal{M}_i, \mathcal{G} | 1-\lambda, \lambda) = [(1-\lambda) \wedge \lambda] \|P_1 - P_2\| .$$

Proof :

This follows readily from section 3 in [19] .

Corollary 4.6.

$$\delta_2(\mathcal{M}_i, \mathcal{G}) = \sup_{\theta_1, \theta_2} \delta(\mathcal{M}_i, \mathcal{G}_{\{\theta_1, \theta_2\}})$$

Proof :

This follows from theorems 4.4 and 4.5.

□

The sublinear function criterion yield :

Proposition 4.7. If Θ is finite then :

$$\delta_{(k)}(\mathcal{M}_i, \mathcal{G}) = \sup_{\gamma \in \Gamma_{(k)}} [\gamma(\mathcal{G}) - \gamma(1, 1, \dots, 1)] .$$

5. Two inequalities.

As stated before, [20], the quantities $\delta(\mathcal{M}_i, \mathcal{M}_a) - \delta(\mathcal{E}, \mathcal{M}_a)$ and $\delta(\mathcal{M}_i, \mathcal{E})$ both provide monotonically increasing and continuous measures of the amount of information in the experiment \mathcal{E} . They also obtain the same extreme values, 0 and $\delta(\mathcal{M}_i, \mathcal{M}_a)$, for, respectively, $\mathcal{E} = \mathcal{M}_i$ and $\mathcal{E} = \mathcal{M}_a$.

If Θ is finite then, by continuity and compactness, convergence of one of these measures towards one of the extreme values entails the convergence of the other towards the same extreme value. In spite of these similarities, the first measure is concave in \mathcal{E} while the second is convex in \mathcal{E} .

The purpose of this section is to establish inequalities which yield numerical estimates of how close one of these measures are to one of the extreme values when the deviation of the other measure from the same extreme value is given. These inequalities are all discouraging trivial when Θ is infinite. We shall for this reason, assume throughout this section that Θ is finite. The experiment will be denoted by $\mathcal{E} = (P_\theta; \theta \in \Theta)$ and we let m denote the number of elements in Θ . Thus $\delta(\mathcal{M}_i, \mathcal{M}_a) = 2 - \frac{2}{m}$.

The quantity $\delta(\mathcal{M}_i, \mathcal{M}_a) - \delta(\mathcal{E}, \mathcal{M}_a)$ may be bounded below by the statistical distances between parameter points as follows :

Proposition 5.1.

$$\frac{2}{m^2} \max_{\theta_1 \neq \theta_2} \delta(\mathcal{M}_i, \mathcal{E}_{\{\theta_1, \theta_2\}}) \leq \delta(\mathcal{M}_i, \mathcal{M}_a) - \delta(\mathcal{E}, \mathcal{M}_a).$$

This yields the first of the promised inequalities :

Corollary 5.2.

$$\frac{1}{m(m-1)} \delta(\mathcal{M}_i, \mathcal{E}) \leq \delta(\mathcal{M}_i, \mathcal{M}_a) - \delta(\mathcal{E}, \mathcal{M}_a) \leq \delta(\mathcal{M}_i, \mathcal{E}) .$$

Proof of the corollary: By the triangle inequality :

$\delta(\mathcal{M}_i, \mathcal{M}_a) \leq \delta(\mathcal{M}_i, \mathcal{E}) + \delta(\mathcal{E}, \mathcal{M}_a)$ and this gives the right hand side inequality. By Corollary 4.2 and the proposition we find :

$$\begin{aligned} \frac{1}{m(m-1)} \delta(\mathcal{M}_i, \mathcal{E}) &\leq \frac{1}{m(m-1)} 2(1 - \frac{1}{m}) \max_{\theta_1 \neq \theta_2} \delta(\mathcal{M}_i, \mathcal{E}_{\{\theta_1, \theta_2\}}) \\ &= \frac{2}{m^2} \max_{\theta_1 \neq \theta_2} \delta(\mathcal{M}_i, \mathcal{E}_{\{\theta_1, \theta_2\}}) \leq 2 - \frac{2}{m} - \delta(\mathcal{E}, \mathcal{M}_a) . \end{aligned}$$

Example 5.3. Let X_0, X_1, X_2, \dots be an aperiodic and irreducible Markov chain with finite state space Θ : Put $\mathcal{E}_n = (P_{\theta, n} : \theta \in \Theta)$ where, for each θ and each positive integer n , $P_{\theta, n}$ is the conditional distribution of X_n given $X_0 = \theta$. Then, as is shown by Lindqvist [16], $\sqrt[n]{\delta(\mathcal{M}_i, \mathcal{E}_n)}$ converges, as $n \rightarrow \infty$, to the largest number in $[0, 1[$ which is a modulus of a characteristic root of the transition matrix.

Hence, by the corollary, $\sqrt[n]{\delta(\mathcal{M}_i, \mathcal{M}_a) - \delta(\mathcal{E}_n, \mathcal{M}_a)}$ converges to the same limit.

Proof of proposition 5.1: Let θ_1 and θ_2 be distinct points in Θ and let $\lambda \in \Lambda$. We shall try to find a lower bound for

$$(\S) \quad \|\vee_{\theta}^{\lambda} P_{\theta}\| - \frac{1}{m^2} \|P_{\theta_1} \vee P_{\theta_2}\|$$

when $\|P_{\theta_1} \vee P_{\theta_2}\| = b+1$ is given. Here b is a , and may be any, number in $[0, 1]$. It follows, by the requirement on $\|P_{\theta_1} \vee P_{\theta_2}\|$, that there is an event A so that

$$P_{\theta_2}(A) + 1 - P_{\theta_1}(A) = b+1 ; \text{ i.e. } P_{\theta_2}(A) = P_{\theta_1}(A) + b .$$

Denote the restriction of P_{θ} to $\{\emptyset, A, A^c, A \cup A^c\}$ by \tilde{P}_{θ} .

Then $\|\tilde{P}_{\theta_1} \vee \tilde{P}_{\theta_2}\| = \|P_{\theta_1} \vee P_{\theta_2}\| = b+1$ while, since restrictions reduce information, $\|\vee_{\theta} \lambda_{\theta} P_{\theta}\| \geq \|\vee_{\theta} \lambda_{\theta} \tilde{P}_{\theta}\|$. It follows that we may, for the purpose of maximizing (§), assume that the sample space is a two point set, say $\{A, A^c\}$. Put $p_{\theta} = P_{\theta}(A)$. Then (§) may be written :

$$(\S\S) \quad \vee_{\theta} \lambda_{\theta} (1 - p_{\theta}) + \vee_{\theta} \lambda_{\theta} p_{\theta} - \frac{b+1}{m^2}$$

where $(1-p_{\theta_1}) \vee (1-p_{\theta_2}) + p_{\theta_1} \vee p_{\theta_2} = 1+b = 1-p_{\theta_1} + p_{\theta_2}$.

Put $U = \{\theta_0, \theta_1\}$, $W = \Theta - U$ and $F = \vee_{\theta} \lambda_{\theta} (1-p_{\theta}) + \vee_{\theta} \lambda_{\theta} p_{\theta}$.

Then :

$$\begin{aligned} F &= [\vee_U \lambda_{\theta} (1-p_{\theta}) \vee \vee_W \lambda_{\theta} (1-p_{\theta})] + [\vee_U \lambda_{\theta} p_{\theta} \vee \vee_W \lambda_{\theta} p_{\theta}] \\ &\geq [\vee_W \lambda_{\theta} (1-p_{\theta}) + \vee_W \lambda_{\theta} p_{\theta}] \vee [\vee_U \lambda_{\theta} (1-p_{\theta}) + \vee_U \lambda_{\theta} p_{\theta}]. \end{aligned}$$

The expression in the first bracket is a monotone function of ξ_W , and is consequently minimized when p_{θ} does not depend on θ when $\theta \in W$. Hence

$$F \geq [\vee_W \lambda_{\theta}] \vee [\vee_U \lambda_{\theta} (1-p_{\theta}) + \vee_U \lambda_{\theta} p_{\theta}].$$

Clearly $\vee_W \lambda_{\theta} \geq \frac{1-\lambda_{\theta_0} - \lambda_{\theta_1}}{m-2}$.

Thus :

$$F \geq \frac{1-\lambda_{\theta_0} - \lambda_{\theta_1}}{m-2} \vee (\lambda_{\theta_0} + \lambda_{\theta_1}) \left[\vee_{\theta \in U} \lambda_{\theta} (1-p_{\theta}) + \vee_{\theta \in U} \lambda_{\theta} p_{\theta} \right]$$

where $(n_{\theta_0}, n_{\theta_1})$ is and may be any probability distribution on $\{\theta_0, \theta_1\}$.

Consider now the dichotomy \mathcal{G}_U i.e. :

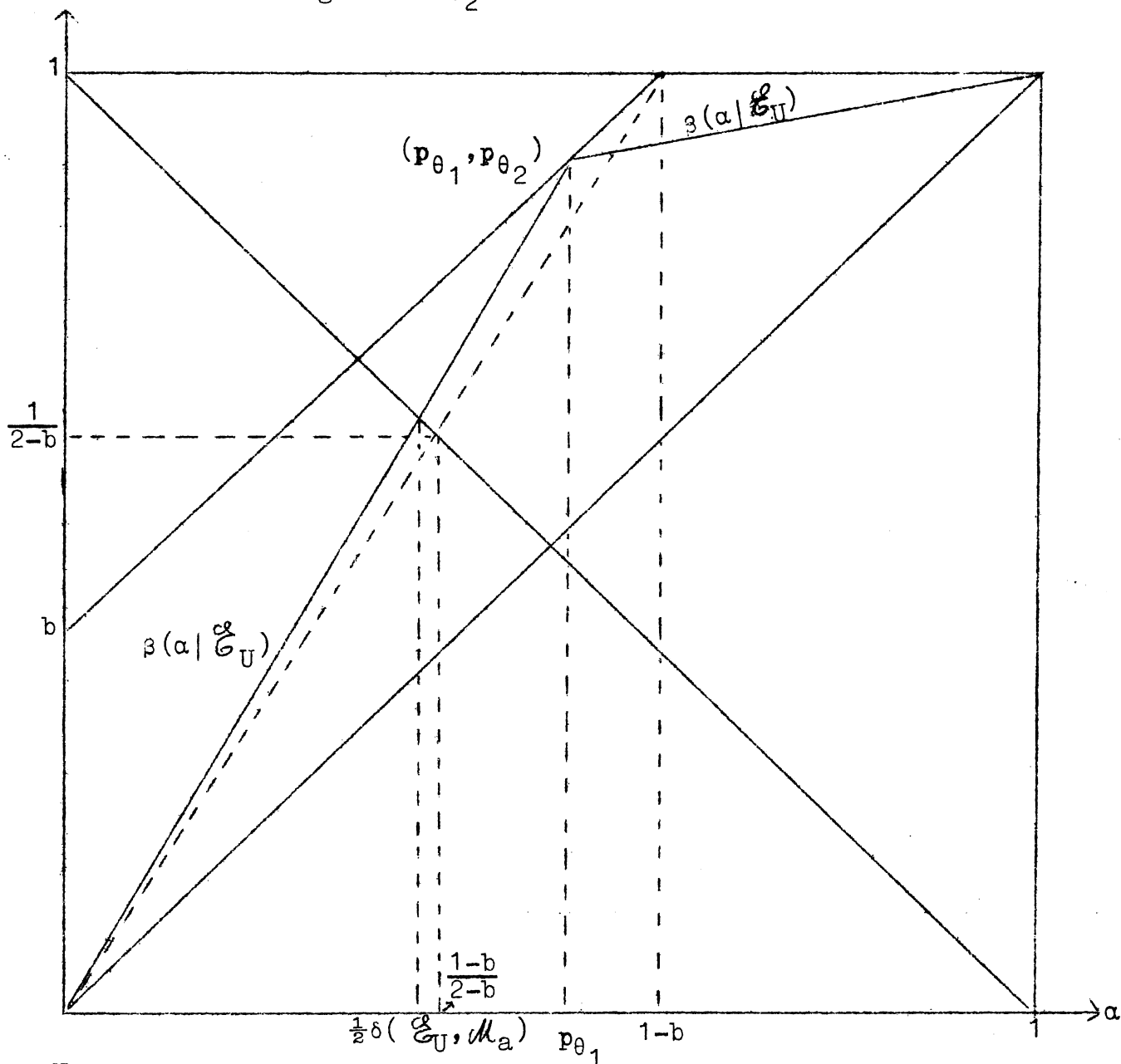
	A^c	A
θ_1	$1-p_{\theta_1}$	p_{θ_1}
θ_2	$1-p_{\theta_2}$	p_{θ_2}

Let $(p_{\theta_1}, p_{\theta_2})$ range through the set of all numbers (x_1, x_2) in $[0,1]^2$ such that $x_2 = x_1 + b$. Which of these dichotomies have the maximum distance to \mathcal{M}_a ?

Utilizing corollary 16 in [19] we find, see fig. **on next page**, that $\delta(\mathcal{G}_U, \mathcal{M}_a)$ obtains its maximum $\frac{2-2b}{2-b}$ when $p_{\theta_1} = 0$ and $p_{\theta_2} = 1-b$.

Here is the power function diagram for \mathcal{E}_U :

$\beta(\alpha | \mathcal{E}_U)$ = maximal power at level α for testing θ_1
against θ_2



Now

$$\delta(\mathcal{E}_U, \mathcal{M}_a) = 2(1 - \inf_{\kappa} [\nu_{\kappa}^U \theta (1 - p_{\theta}) + \nu_{\kappa}^U \theta p_{\theta}]) .$$

Thus :

$$2 \frac{1-b}{2-b} = \max_{\mathcal{E}_U} \delta(\mathcal{E}_U, \mathcal{M}_a) = 2(1 - \min_{\mathcal{E}_U} \min_{\kappa} [\nu_{\kappa}^U \theta (1 - p_{\theta}) + \nu_{\kappa}^U \theta p_{\theta}])$$

so that:

$$\min_{\mathcal{E}_U} \min_{\kappa} [\sum_U \kappa_{\theta} (1-p_{\theta}) + \sum_U \kappa_{\theta} p_{\theta}] = \frac{1}{2-b} .$$

Hence :

$$F \geq \left[\frac{1-\lambda_{\theta_0}-\lambda_{\theta_1}}{m-2} \right] \sum_U [(\lambda_{\theta_0} + \lambda_{\theta_1}) \frac{1}{2-b}] \quad \text{and the expression}$$

on the right is minimized when

$$\frac{\lambda_{\theta_0} + \lambda_{\theta_1}}{2-b} = \frac{1-\lambda_{\theta_0}-\lambda_{\theta_1}}{m-2} = \frac{1}{m-b}$$

$$\text{Thus } F \geq \frac{1}{m-b} .$$

It follows that

$$\begin{aligned} & \left\| \sum_{\theta} \lambda_{\theta} P_{\theta} \right\| - \frac{H}{2} \left\| P_{\theta_1} \vee P_{\theta_2} \right\| \\ & \geq \frac{1}{m-b} - \frac{b+1}{m^2} = \frac{1}{m-b} - \frac{H}{2}(b+1) \geq \frac{1}{m} - \frac{H}{2} \quad \text{where } H = \frac{2}{m^2} . \end{aligned}$$

Now :

$$\delta(\mathcal{M}_i, \mathcal{E}_U) = \frac{1}{2} \|P_{\theta_1} - P_{\theta_2}\| = \|P_{\theta_1} \vee P_{\theta_2}\| - 1 .$$

Thus :

$$\left\| \sum_{\theta} \lambda_{\theta} P_{\theta} \right\| - \frac{H}{2} (\delta(\mathcal{M}_i, \mathcal{E}_U) + 1) \geq \frac{1}{m} - \frac{H}{2}$$

or

$$\frac{H}{2} \delta(\mathcal{M}_i, \mathcal{E}_U) + 1 - \left\| \sum_{\theta} \lambda_{\theta} P_{\theta} \right\| \leq 1 - \frac{1}{m} .$$

Maximizing the left hand side w.r.t. λ we find , using corollary 3.2 , that :

$$H \delta(\mathcal{M}_i, \mathcal{E}_U) + 2 \delta(\mathcal{E}, \mathcal{M}_a) \leq 2 - \frac{2}{m}$$

so that :

$$H \max_U \delta(\mathcal{M}_i, \mathcal{E}_U) + 2\delta(\mathcal{E}, \mathcal{M}_a) \leq 2 - \frac{2}{m}$$

where U ranges through all two point sub sets of Θ .

□

An upper bound for $\delta(\mathcal{M}_i, \mathcal{E})$ in terms of the pairwise deficiencies $\delta(\mathcal{M}_i, \mathcal{E}_{\{\theta_1, \theta_2\}})$ is given by :

Proposition 5.4.

$$\frac{2}{m(m-1)} (1 - \min_{\theta_1 \neq \theta_2} \delta(\mathcal{M}_i, \mathcal{E}_{\{\theta_1, \theta_2\}})) \leq 2 - \frac{2}{m} - \delta(\mathcal{M}_i, \mathcal{E}) .$$

Corollary 5.5.

$$\frac{1}{m(m-1)^2} \delta(\mathcal{E}, \mathcal{M}_a) \leq 2 - \frac{2}{m} - \delta(\mathcal{M}_i, \mathcal{E}) \leq \delta(\mathcal{E}, \mathcal{M}_a) .$$

Proof :

The right most inequality is the same as the right most inequality of corollary 5.2 . Let θ_1 and θ_2 be distinct points in Θ and let $(\lambda_{\theta_1}, \lambda_{\theta_2})$ run through all probability distributions on $\{\theta_1, \theta_2\}$. Then, by corollary 3.2 and proposition 5.4 :

$$\delta(\mathcal{E}_{\{\theta_1, \theta_2\}}, \mathcal{M}_a) = 2 \sup b(\lambda_{\theta_1}, \lambda_{\theta_2} | \mathcal{E}_{\{\theta_1, \theta_2\}}) =$$

$$\begin{aligned}
&= 2 \sup \|\lambda_{\theta_1} P_{\theta_1} \wedge \lambda_{\theta_2} P_{\theta_2}\| \leq 2 \|P_{\theta_1} \wedge P_{\theta_2}\| = 2[1 - \frac{1}{2} \|P_{\theta_1} - P_{\theta_2}\|] = \\
&= 2(1 - \delta(\mathcal{M}_i, \mathcal{E}_{\{\theta_1, \theta_2\}})) \leq m(m-1)[2 - \frac{2}{m} - \delta(\mathcal{M}_i, \mathcal{E})] . \quad \text{Thus, by}
\end{aligned}$$

corollary 3.9 :

$$\frac{1}{m(m-1)^2} \delta(\mathcal{E}, \mathcal{M}_a) \leq \max_{\theta_1 \neq \theta_2} \frac{1}{m(m-1)} \delta(\mathcal{E}_{\{\theta_1, \theta_2\}}, \mathcal{M}_a) \leq 2 - \frac{2}{m} - \delta(\mathcal{M}_i, \mathcal{E}) .$$

□

Proof of the proposition :

The inequality may, by theorem 4.5 be written :

$$\frac{2}{m(m-1)} \|P_{\theta_1} \wedge P_{\theta_2}\| + \delta(\mathcal{M}_i, \mathcal{E}) \leq 2 - \frac{2}{m} , \quad \theta_1 \neq \theta_2 .$$

Let θ_1 and θ_2 be distinct points in Θ and consider the problem of maximizing $\delta(\mathcal{M}_i, \mathcal{E})$ when $\|P_{\theta_1} \wedge P_{\theta_2}\| = \xi$ is given. Clearly ξ is, and may be, any number in $[0, 1]$. We may, without loss of generality assume that \mathcal{E} is a standard experiment with standard measure S on the simplex K of all prior distributions on Θ . Thus we have to maximize $\delta(\mathcal{M}_i, S)$ when $\int (y_{\theta_1} \wedge y_{\theta_2}) S(dy) = \xi$ is given.

Let D be any dilatation on K such that

$$D(\{x: x_{\theta_1} \geq x_{\theta_2}\} | y) = 1 \quad \text{if} \quad y_{\theta_1} \geq y_{\theta_2}$$

and

$$D(\{x: x_{\theta_1} \leq x_{\theta_2}\} | y) = 1 \quad \text{if} \quad y_{\theta_1} \leq y_{\theta_2} .$$

Then, since DS is more informative than S :

$\delta(\mathcal{M}_i, DS) \geq \delta(\mathcal{M}_i, S)$. On the other hand, by the particular properties of D :

$$\int (x_{\theta_1} \wedge x_{\theta_2}) D(dx|y) = y_{\theta_1} \wedge y_{\theta_2}$$

so that

$$\int (x_{\theta_1} \wedge x_{\theta_2}) DS(dx) = \int (y_{\theta_1} \wedge y_{\theta_2}) S(dy) .$$

Put $K_1 = \{x: x \in K, x_{\theta_1} \geq x_{\theta_2}\}$ and $K_2 = \{x: x \in K, x_{\theta_2} \geq x_{\theta_1}\}$.

It is easily seen * that the extreme points of K_1 are e^θ ; $\theta \neq \theta_2$ and $\frac{1}{2}(e^{\theta_1} + e^{\theta_2})$ while the extreme points of K_2 are e^θ ; $\theta \neq \theta_1$ and $\frac{1}{2}(e^{\theta_1} + e^{\theta_2})$. Define for each $y \in K_1$ the probability measure $D(\cdot|y)$ on K_1 by :

$$D(e^{\theta_1}|y) = y_{\theta_1} - y_{\theta_2}$$

$$D(\frac{1}{2}(e^{\theta_1} + e^{\theta_2})|y) = 2y_{\theta_2} \quad \text{and}$$

$$D(e^\theta|y) = y_\theta \quad \text{if } \theta \neq \theta_1 \text{ and } \theta \neq \theta_2 .$$

Similarly, if $y \in K_2$ we define the probability measure $D(\cdot|y)$ on K_2 by

*) e^θ is the θ -th unit vector in R^Θ i.e.
 $e^\theta = (0, 0, \dots, \overset{\theta}{1}, \dots, 0)$.

$$D(e^{\theta_2} | y) = y_{\theta_2} - y_{\theta_1}$$

$$D(\frac{1}{2}(e^{\theta_1} + e^{\theta_2}) | y) = 2y_{\theta_1}$$

and

$$D(e^{\theta} | y) = y_{\theta} \quad \text{if } \theta \neq \theta_1 \text{ and } \theta \neq \theta_2$$

Obviously there is no conflict between these definitions when $y \in K_1 \cap K_2$. It is easily checked that D is a dilatation preserving the regions K_1 and K_2 . This dilatation carries any standard measure into a standard measure concentrated on the $(m+1)$ point set $\{e^{\theta}; \theta \in \Theta\} \cup \{\frac{1}{2}(e^{\theta_1} + e^{\theta_2})\}$. The only experiment satisfying this condition is * the experiment \mathcal{E}_{η} where $\eta_{\theta} = \xi$ or $= 0$ as $\theta \in \{\theta_1, \theta_2\}$ or $\theta \notin \{\theta_1, \theta_2\}$. Thus $\delta(\mathcal{M}_i, \mathcal{E}_{\eta})$ is the desired maximum. We shall satisfy ourselves with an upper bound for this quantity. Suppose first that $\xi \leq \frac{1}{m-1}$. Approximate each P_{θ} with the distribution Q which assigns masses $\frac{1-(m-1)\xi}{m}$ to θ_1 and θ_2 , masses $\frac{1+\xi}{m}$ to each of the other points in Θ , and mass ξ to Θ itself. Note that Θ is, at the same time, both the parameter set of \mathcal{E}_{η} and a point in the sample space of \mathcal{E}_{η} . Simple calculations yield :

$$\|P_{\theta} - Q\| \leq 2(1 - \frac{1}{m} - \frac{\xi}{m}) \quad ; \quad \theta \in \Theta.$$

*) The experiments \mathcal{E}_{η} ; $\eta \in [0, 1]^{\Theta}$ are defined in example 3.12.

Thus

$$(\S) \quad \delta(\mathcal{M}_i, \mathcal{E}_\eta) \leq 2 - \frac{2}{m} - \frac{2\xi}{m} \quad \text{when } \xi \leq \frac{1}{m-1}.$$

Suppose next that $\xi \geq \frac{1}{m-1}$. Then, by example 3.12 ,
 $\delta(\mathcal{M}_i, \mathcal{E}_\eta) \leq \delta(\mathcal{M}_i, \mathcal{E}_\rho)$ where $\rho_\theta = \frac{1}{m-1}$ or $= 0$ as
 $\theta \in \{\theta_1, \theta_2\}$ or $\theta \notin \{\theta_1, \theta_2\}$. Hence, by (\S) :

$$\delta(\mathcal{M}_i, \mathcal{E}_\eta) \leq 2 - \frac{2}{m} - \frac{2}{m} \frac{2}{m-1} \leq 2 - \frac{2}{m} - \kappa \xi$$

where $\kappa = \frac{2}{m(m-1)}$. Hence, by (\S) again :

$\delta(\mathcal{M}_i, \mathcal{E}_\eta) \leq 2 - \frac{2}{m} - \kappa \xi$ for all ξ and this is the
 desired inequality.

□

6. Replicated dichotomies.

We shall in this section assume that $\Theta = \{1,2\}$ i.e. that our experiments are dichotomies. The Baye's risk function $\lambda \rightsquigarrow b(1-\lambda, \lambda | \mathcal{E})$ may then be written :

$$b(1-\lambda, \lambda | \mathcal{E}) = 1 - \| (1-\lambda)P_1 \vee \lambda P_2 \| = \| (1-\lambda)P_1 \wedge \lambda P_2 \| .$$

Hence

$$0 \leq b(1-\lambda, \lambda | \mathcal{E}) \leq (1-\lambda) \wedge \lambda \quad ; \quad \lambda \in [0,1] .$$

Note that the right hand side is $b(1-\lambda, \lambda | \mathcal{M}_1)$ while the left hand side is $b(1-\lambda, \lambda | \mathcal{M}_a)$. Conversely, any concave function f on $[0,1]$ such that $0 \leq f(\lambda) \leq (1-\lambda) \wedge \lambda \quad ; \quad \lambda \in [0,1]$ is of the form $f(\lambda) = b(1-\lambda, \lambda | \mathcal{E})$ where \mathcal{E} is, up to equivalence, determined by f . A particularly interesting aspect of this representation is the relation

$$b(1-\lambda, \lambda | \sup_t \mathcal{E}_t) = \inf_t b(1-\lambda, \lambda | \mathcal{E}_t)$$

which is valid for any family $\{\mathcal{E}_t; t \in T\}$ of dichotomies.

Deficiencies are easily described in terms of these functions.

By [19] and [22] :

$$\delta(\mathcal{E}, \mathcal{F}) = 2 \sup_{\lambda} [b(1-\lambda, \lambda | \mathcal{E}) - b(1-\lambda, \lambda | \mathcal{F})]$$

for any pair $(\mathcal{E}, \mathcal{F})$ of dichotomies.

We refer to [19] and [22] for this and various other results on dichotomies.

There is a simple connection between the Hellinger transform and the Baye's risk for dichotomies :

Theorem 6.1.

$$H(t_1, t_2 | \mathcal{G}) / t_1 t_2 = \int_0^1 \frac{b(1-\lambda, \lambda | \mathcal{G})}{(1-\lambda)^{t_1+1} \lambda^{t_2+1}} d\lambda = \int_0^\infty \frac{\|xP_1 \wedge P_2\|}{x^{t_2+1}} dx$$

The last equality follows by substituting $x = \frac{1-\lambda}{\lambda}$ in the middle integral. As all three expressions define continuous and affine functionals it suffices, by section 4 in [19], to establish the identities for double dichotomies - and in that case they follow by a simple integration.

As mentioned above a non negative and concave function on $[0,1]$, is a minimum Baye's risk function for the decision problem defined in section 3 if and only if it is dominated by the triangular function $\lambda \rightsquigarrow \lambda \wedge (1-\lambda)$. Thus the functions $\lambda \rightsquigarrow \frac{(1-\lambda)^{1-t} \lambda^t}{(1-t)^t}$; $t \in]0,1[$ are not minimum Baye's risk functions. We may, however, consider the function $\lambda \rightsquigarrow \frac{(1-\lambda)^{1-t} \lambda^t}{(1-t)^t}$, for $t \in]0,1[$, as the minimum Baye's risk functions in a "pseudo" dichotomy (U_t, V_t) . Here U_t and V_t are the measures on $[0,\infty[$ whose densities w.r.t. Lebesgue measure are, respectively, $x \rightsquigarrow x^{-t-1}$ and $x \rightsquigarrow x^{-t}$.

Thus :

Proposition 6.2.

$$\frac{(1-\lambda)^{1-t} \lambda^t}{(1-t)t} = \|(1-\lambda)U_t \wedge \lambda V_t\| = \int_0^\infty \frac{(1-\lambda) \wedge \lambda x}{x^{t+1}} dx ;$$

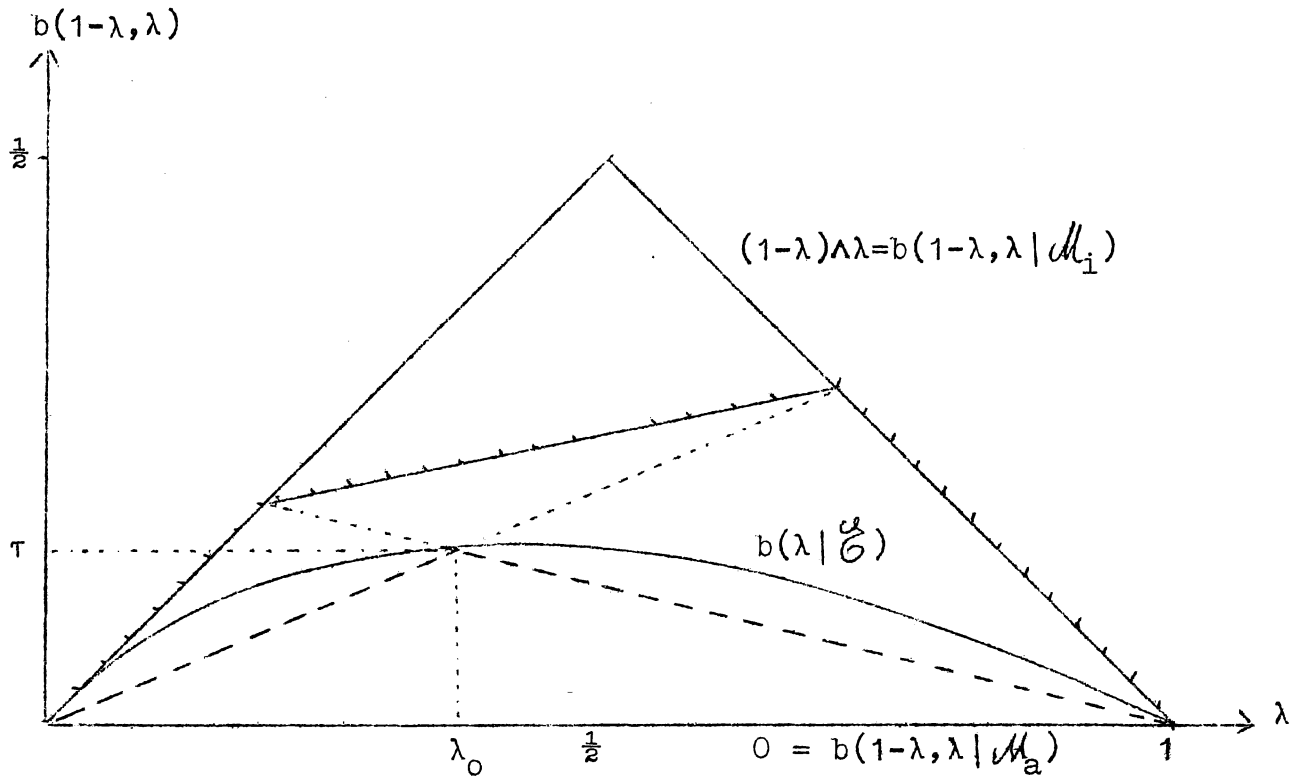
or equivalently :

$$\frac{\lambda_1^{t_1} \lambda_2^{t_2}}{t_1 t_2} = \int_0^1 \frac{\lambda_1(1-y) \wedge \lambda_2 y}{(1-y)^{t_1+1} y^{t_2+1}} dy$$

for each prior distribution (λ_1, λ_2) and each non degenerate prior distribution (t_1, t_2) .

The connection between Hellinger transforms and minimum Baye's risk functions follows readily from the last identity of proposition 6.2 by integrating both sides w.r.t. the standard measure of \mathcal{G} .

As an application consider the problem of finding bounds for $H(1-t, t | \mathcal{G})$ in terms of $b(1-\lambda, \lambda | \mathcal{G})$. Put $\lambda = \lambda_0$ and suppose $b(1-\lambda_0, \lambda_0 | \mathcal{G}) = \tau$ is given. Here is a possible graph of $\lambda \rightsquigarrow b(1-\lambda, \lambda | \mathcal{G})$:



Legend : - - - - - Graph of $\lambda \rightsquigarrow b(1-\lambda, \lambda | \mathcal{E}_0)$
 ↗ ↘ ↗ ↘ ↗ ↘ Graph of $\lambda \rightsquigarrow b(1-\lambda, \lambda | \mathcal{E}_1)$

By convexity : $\mathcal{E}_1 \leq \mathcal{E} \leq \mathcal{E}_0$ where the graph of $b(1-\lambda, \lambda | \mathcal{E}_0)$ is the triangle joining $(0,0)$, (λ_0, τ) and $(1,0)$ while the graph of $b(1-\lambda, \lambda | \mathcal{E}_1)$ is the quadrangle joining $(0,0)$, $((1-\lambda_0+\tau)^{-1}\tau, (1-\lambda_0+\tau)^{-1}\tau)$, $((\lambda_0+\tau)^{-1}\lambda_0, (\lambda_0+\tau)^{-1}\tau)$ and $(1,0)$.

Hence, since $x \rightsquigarrow x^{1-t}x^t$ is concave :

$$H(1-t, t | \mathcal{E}_0) \leq H(1-t, t | \mathcal{E}) \leq H(1-t, t | \mathcal{E}_1) .$$

It is easily checked that \mathcal{E}_0 and \mathcal{E}_1 may be represented as, respectively, the dichotomies :

$\theta \backslash x$	0	1	2
1	$1-\xi$	0	ξ
2	0	$1-\eta$	η

and

$\theta \backslash x$	0	1
1	$1-p$	p
2	$1-q$	q

where $\xi = \frac{\tau}{1-\lambda_0}$, $\eta = \frac{\tau}{\lambda_0}$, $p = \frac{\tau(\lambda_0 - \tau)}{\lambda_0(1-\lambda_0) - \tau^2}$ and

$$q = \frac{(1-\lambda_0)(\lambda_0 - \tau)}{\lambda_0(1-\lambda_0) - \tau^2} .$$

It follows that :

$$H(1-t, t | \mathcal{G}_0) = \frac{\tau}{(1-\lambda_0)^{1-t} \lambda_0^t}$$

and

$$H(1-t, t | \mathcal{G}_1) = \frac{\lambda_0^{1-t}(1-\lambda_0-\tau)\tau^t + \tau^{1-t}(\lambda_0-\tau)(1-\lambda_0)^t}{\lambda_0(1-\lambda_0) - \tau^2} .$$

We have proved :

Theorem 6.3.

Let λ and t be non degenerate prior distributions.

Then :

$$\begin{aligned} \frac{b(\lambda | \mathcal{G})}{\lambda_1^{t_1} \lambda_2^{t_2}} &\leq H(t | \mathcal{G}) \leq [\lambda_1 \lambda_2^{-b(\lambda | \mathcal{G})}]^{-1} \{ [\lambda_1^{-b(\lambda | \mathcal{G})}] \lambda_2^{t_1 b(\lambda | \mathcal{G}) t_2} + \\ &+ [\lambda_2^{-b(\lambda | \mathcal{G})}] \lambda_1^{t_2 b(\lambda | \mathcal{G}) t_1} \} . \end{aligned}$$

Remark 1.

The left " \leq " imply that :

$$b(\lambda | \mathcal{G}) \leq (\lambda_1 \vee \lambda_2) \inf_t H(t | \mathcal{G})$$

so that

$$\delta(\mathcal{G}, \mathcal{M}_a)/2 \leq \inf_t H(t | \mathcal{G}) .$$

Remark 2.

Let f_1 and f_2 be densities of, respectively, P_1 and P_2 w.r.t. $P_1 + P_2$. Then the left " \leq " may be proved as follows :

$$\begin{aligned} b(\lambda | \mathcal{G}) &= \int \lambda_1 f_1 \wedge \lambda_2 f_2 = \int (\lambda_1 f_1 \wedge \lambda_2 f_2)^{t_1} (\lambda_1 f_1 \wedge \lambda_2 f_2)^{t_2} \\ &\leq \int (\lambda_1 f_1)^{t_1} (\lambda_2 f_2)^{t_2} = \lambda_1^{t_1} \lambda_2^{t_2} H(t | \mathcal{G}) \end{aligned}$$

Corollary 6.4.

Let t be a non degenerate prior distribution. Then :

$$\|P_1 \wedge P_2\| \leq H(t | \mathcal{G}) \leq [1 + \|P_1 \wedge P_2\|]^{-1} [\|P_1 \wedge P_2\|^{t_1} + \|P_1 \wedge P_2\|^{t_2}]$$

Proof :

$$\text{Put } \lambda_1 = \lambda_2 = \frac{1}{2} .$$

□

Corollary 6.5.

Let λ be a non degenerate prior distribution. Then

$$\begin{aligned} \frac{b(\lambda | \mathcal{G})}{\sqrt{\lambda_1 \lambda_2}} &\leq \gamma(\mathcal{G}) \leq [\lambda_1 \lambda_2^{-b(\lambda | \mathcal{G})}]^{-1} \sqrt{b(\lambda | \mathcal{G})} [\sqrt{\lambda_2} (\lambda_1^{-b(\lambda | \mathcal{G})} + \\ &\quad + \sqrt{\lambda_1} (\lambda_2^{-b(\lambda | \mathcal{G})}))] \end{aligned}$$

where $\gamma(\mathcal{G}) = H(\frac{1}{2}, \frac{1}{2} | \mathcal{G})$ is the affinity between P_1 and P_2 .

Proof :

$$\text{Put } t_1 = t_2 = \frac{1}{2}.$$

□

Corollary 6.6.

$$\|P_1 \wedge P_2\| \leq \gamma(\mathcal{G}) \leq 2\sqrt{\|P_1 \wedge P_2\|} [1 + \|P_1 \wedge P_2\|]^{-1} \leq 2\sqrt{\|P_1 \wedge P_2\|}$$

Proof :

$$\text{Put } t_1 = t_2 = \frac{1}{2} \text{ in corollary 6.4.}$$

□

Remark :

The left most inequality may, since, $\|P_1 \wedge P_2\| = 1 - \frac{1}{2}\|P_1 - P_2\|$ be written :

$$\|P_1 - P_2\| \geq D(P_1, P_2)^2$$

where $D(P_1, P_2)$ is the Hellinger distance between P_1 and P_2 .
A simpler derivation of this inequality may be found in Le Cam [13].

The upper bound, $H(t_1, t_2 | \mathcal{G}_1)$, for $H(t_1, t_2 | \mathcal{G})$ may be considerably lowered if, in addition to $b(1-\lambda_0, \lambda_0 | \mathcal{G})$, we are given a derivative of $\lambda \mapsto b(1-\lambda, \lambda | \mathcal{G})$ at $\lambda = \lambda_0$.

Let, for each $\lambda \in]0, 1[$, $b^*(1-\lambda, \lambda | \mathcal{G})$ be any number between the left and the right derivative of $\lambda \mapsto b(1-\lambda, \lambda | \mathcal{G})$ at λ .
If $\lambda = 0$ or $\lambda = 1$ then $b^*(1, 0 | \mathcal{G})$ and $b^*(0, 1 | \mathcal{G})$ denote respectively, the right derivative in $\lambda = 0$ and the left derivative in $\lambda = 1$.

Then we have :

Theorem 6.7.

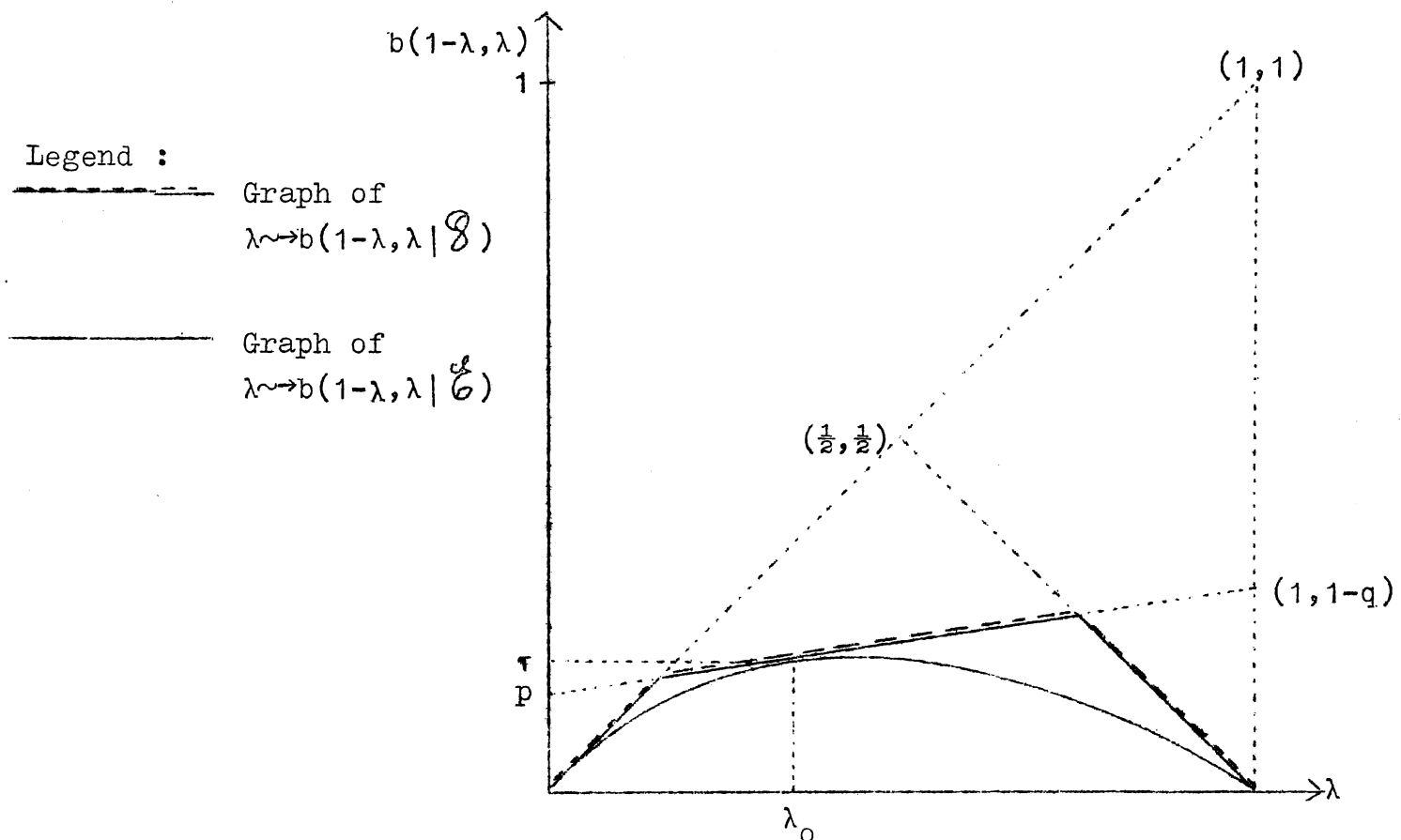
For any pair (λ, t) of prior distributions $\lambda = (\lambda_1, \lambda_2)$ and $t = (t_1, t_2)$:

$$H(t|\mathcal{G}) \leq [1-b(\lambda|\mathcal{G}) + \lambda_2 b^*(\lambda|\mathcal{G})]^{t_1} [b(\lambda|\mathcal{G}) + \lambda_1 b^*(\lambda|\mathcal{G})]^{t_2} \\ + [b(\lambda|\mathcal{G}) - \lambda_2 b^*(\lambda|\mathcal{G})]^{t_1} [1-b(\lambda|\mathcal{G}) - \lambda_1 b^*(\lambda|\mathcal{G})]^{t_2} .$$

Proof :

Fix $\lambda = (1-\lambda_0, \lambda_0)$ and put $\tau = b(1-\lambda_0, \lambda_0|\mathcal{G})$,
 $\kappa = b^*(1-\lambda_0, \lambda_0|\mathcal{G})$, $p = \tau - \kappa\lambda_0$ and $q = 1-\tau-\kappa(1-\lambda_0)$.

Here is a picture of the situation :



The line $\lambda \rightsquigarrow \tau + \kappa(\lambda - \lambda_0)$ is clearly tangent to the graph of $\lambda \rightarrow b(1-\lambda, \lambda | \mathcal{E})$ at $\lambda = \lambda_0$. Hence, since $b(1-\lambda, \lambda | \mathcal{E})$ is concave in λ , $\tau + \kappa(\lambda - \lambda_0) \geq b(1-\lambda, \lambda | \mathcal{E})$ or equivalently :

$$\mathcal{G} \leq \mathcal{E} \quad \text{where} \quad b(1-\lambda, \lambda | \mathcal{G}) = [\tau + \kappa(\lambda - \lambda_0)] \wedge \lambda \wedge (1-\lambda) .$$

The dichotomy \mathcal{G} is the double dichotomy.

$\theta \backslash x$	0	1
1	1-p	p
2	1-q	q

It follows that :

$$H(t | \mathcal{E}) \leq H(t | \mathcal{G}) = (1-p)^{t_1} (1-q)^{t_2} + p^{t_1} q^{t_2} \quad \text{and this is the}$$

desired inequality. □

Corollary 6.8.

$$\begin{aligned} 2H(t | \mathcal{E}) &\leq [\|P_1 \vee P_2\| + b^*(\tfrac{1}{2}, \tfrac{1}{2} | \mathcal{E})]^{t_1} [\|P_1 \wedge P_2\| + b^*(\tfrac{1}{2}, \tfrac{1}{2} | \mathcal{E})]^{t_2} \\ &\quad + [\|P_1 \wedge P_2\| - b^*(\tfrac{1}{2}, \tfrac{1}{2} | \mathcal{E})]^{t_1} [\|P_1 \vee P_2\| - b^*(\tfrac{1}{2}, \tfrac{1}{2} | \mathcal{E})]^{t_2} . \end{aligned}$$

Proof :

Put $\lambda_1 = \lambda_2 = \frac{1}{2}$ and use the identities :

$$\|P_1 \vee P_2\| = 1 + \tfrac{1}{2}\|P_1 - P_2\| \quad , \quad \|P_1 \wedge P_2\| = 1 - \tfrac{1}{2}\|P_1 - P_2\| .$$

□

Corollary 6.9.

$$\gamma(\mathcal{G}) \leq \sqrt{[1-b(\lambda|\mathcal{G})+\lambda_2 b^*(\lambda|\mathcal{G})][b(\lambda|\mathcal{G})+\lambda_1 b^*(\lambda|\mathcal{G})]} \\ + \sqrt{[b(\lambda|\mathcal{G})-\lambda_2 b^*(\lambda|\mathcal{G})][1-b(\lambda|\mathcal{G})-\lambda_1 b^*(\lambda|\mathcal{G})]} t_2$$

Proof : Put $t_1 = t_2 = \frac{1}{2}$.

□

Corollary 6.10.

$$\gamma(\mathcal{G}) \leq \frac{1}{2} \left\{ \sqrt{[\|P_1 \vee P_2\| + b^*(\frac{1}{2}, \frac{1}{2}|\mathcal{G})][\|P_1 \wedge P_2\| + b^*(\frac{1}{2}, \frac{1}{2}|\mathcal{G})]} \right. \\ \left. + \sqrt{[\|P_1 \vee P_2\| - b^*(\frac{1}{2}, \frac{1}{2}|\mathcal{G})][\|P_1 \wedge P_2\| - b^*(\frac{1}{2}, \frac{1}{2}|\mathcal{G})]} \right\} \\ \leq \sqrt{\|P_1 \vee P_2\| \|P_1 \wedge P_2\|}.$$

Remark :

The inequality :

$$\gamma(\mathcal{G}) \leq \sqrt{\|P_1 \vee P_2\| \|P_1 \wedge P_2\|} \quad \text{may be written :}$$

$$\|P_1 - P_2\| \leq \sqrt{2(1+\gamma(\mathcal{G}))} D(P_1, P_2)$$

where $D(P_1, P_2) = \sqrt{2(1-\gamma(\mathcal{G}))}$ is the Hellinger distance between P_1 and P_2 . It follows, since $\gamma(\mathcal{G}) \leq 1$, that

$$\|P_1 - P_2\| \leq 2 D(P_1, P_2).$$

This follows, see Le Cam [13], simply and directly from Schwartz inequality. This inequality, together with the inequality in the remark after corollary 6.6 shows that the statistical distance :

$(P_1, P_2) \rightsquigarrow \|P_1 - P_2\|$, and the Hellinger distance :

$(P_1, P_2) \rightsquigarrow \left[\int (\sqrt{dP_1} - \sqrt{dP_2})^2 \right]^{\frac{1}{2}}$ define the same uniformities.

Proof of the corollary :

The first inequality follows simply by putting $t_1 = t_2 = \frac{1}{2}$ in corollary 6.8. Put
 $F(x) = \sqrt{(B+x)(A+x)} + \sqrt{(B-x)(A-x)}$; $|x| \leq A$ where $A = \|P_1 \wedge P_2\|$
 and $B = \|P_1 \vee P_2\|$. Then F is symmetric and concave on $[-A, A]$.
 It follows that $F(x) \leq F(0) = 2\sqrt{AB}$. □

Corollary 6.11.

For any prior distribution (t_1, t_2) :

$$\begin{aligned} H(t_1, t_2 | \mathcal{E}) &\leq [1 - \tfrac{1}{2}\delta(\mathcal{E}, \mathcal{M}_a)]^{t_1} [\tfrac{1}{2}\delta(\mathcal{E}, \mathcal{M}_a)]^{t_2} \\ &\quad + [\tfrac{1}{2}\delta(\mathcal{E}, \mathcal{M}_a)]^{t_1} [1 - \tfrac{1}{2}\delta(\mathcal{E}, \mathcal{M}_a)]^{t_2} \\ &\leq [\tfrac{1}{2}\delta(\mathcal{E}, \mathcal{M}_a)]^{t_1} + [\tfrac{1}{2}\delta(\mathcal{E}, \mathcal{M}_a)]^{t_2} \end{aligned}$$

Remark :

$t_1 = t_2 = \frac{1}{2}$ yield : $\delta(\mathcal{E}, \mathcal{M}_a) \geq \gamma(\mathcal{E})^2/2$.

Proof :

Let λ^0 be least favorable i.e. $b(\lambda^0 | \mathcal{E}) \geq b(\lambda | \mathcal{E})$ for all $\lambda \in [0, 1]$. We may then take $b^*(\lambda^0 | \mathcal{E}) = 0$. The inequality follows now by inserting $\lambda = \lambda^0$ in the inequality in theorem 6.7 and using corollary 3.2. □

Another interesting inequality is :

Proposition 6.12.

If \mathcal{G} and \mathcal{F} are dichotomies, then

$$b(\lambda | \mathcal{G} \times \mathcal{F}) \geq 2b(\lambda | \mathcal{G})b(\tfrac{1}{2}, \tfrac{1}{2} | \mathcal{F}) .$$

Proof :

This follows from Fubini's theorem and the inequality :

$$\lambda_1 f_1 g_1 \wedge \lambda_2 f_2 g_2 \geq [\lambda_1 f_1 \wedge \lambda_2 f_2](g_1 \wedge g_2)$$

which is valid for any non negative numbers $\lambda_1, \lambda_2, f_1, f_2, g_1$ and g_2 . □

Corollary 6.13.

If \mathcal{G} and \mathcal{F} are dichotomies then :

$$b(\lambda | \mathcal{G} \times \mathcal{F}) \geq b(\lambda | \mathcal{G})b(\mu | \mathcal{F})$$

for all prior distributions λ and μ .

Proof :

$$2b(\tfrac{1}{2}, \tfrac{1}{2} | \mathcal{F}) \geq b(\mu | \mathcal{F}) .$$
□

Corollary 6.14.

If \mathcal{G} and \mathcal{F} are dichotomies then :

$$\delta(\mathcal{G} \times \mathcal{F} | \mathcal{M}_a)/2 \geq [\delta(\mathcal{G}, \mathcal{M}_a)/2] 2b(\tfrac{1}{2}, \tfrac{1}{2} | \mathcal{F}) .$$

Proof :

6.13

This follows from corollary 3.2.

□

Corollary 6.15.

If \mathcal{E} and \mathcal{F} are dichotomies then :

$$\delta(\mathcal{E} \times \mathcal{F} | \mathcal{M}_a)/2 \geq [\delta(\mathcal{E}, \mathcal{M}_a)/2][\delta(\mathcal{F}, \mathcal{M}_a)/2] .$$

Proof :

$$2b(\frac{1}{2}, \frac{1}{2} | \mathcal{F}) \geq \delta(\mathcal{F}, \mathcal{M}_a)/2 .$$

□

By corollaries 6.13 and 6.15 :

$$b(\lambda_1, \lambda_2 | \mathcal{E}^{m+n}) \geq b(\lambda_1, \lambda_2 | \mathcal{E}^m) b(\lambda_1, \lambda_2 | \mathcal{E}^n)$$

and

$$\frac{\delta(\mathcal{E}^{m+n} | \mathcal{M}_a)}{2} \geq \frac{\delta(\mathcal{E}^m | \mathcal{M}_a)}{2} \cdot \frac{\delta(\mathcal{E}^n | \mathcal{M}_a)}{2}$$

It follows that

$\sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)}$; $n=1,2,\dots$ and $\sqrt[n]{b(\lambda | \mathcal{E}^n)}$; $n=1,2,\dots$
converges, as $n \rightarrow \infty$, to, respectively: $\sup_n \sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)/2}$ and
 $\sup_n \sqrt[n]{b(\lambda | \mathcal{E}^n)}$. Furthermore, since

$$b(\lambda | \mathcal{E}^n) \leq \frac{\delta(\mathcal{E}^n, \mathcal{M}_a)}{2} \leq \frac{1}{\lambda_1 \wedge \lambda_2} b(\lambda | \mathcal{E}^n)$$

when λ is non degenerate, these limits are the same. Thus :

Theorem 6.16.

There is for each dichotomy \mathcal{E} , a constant* $c(\mathcal{E})$ in $[0,1]$
such that, for each non degenerate prior distribution λ :

*) It follows directly from theorem 6.16 that $c(\mathcal{E}^r) = c(\mathcal{E})^r$;
 $r=1,2,\dots$. Also, by this theorem and corollary 6.15 :
 $c(\mathcal{E} \times \mathcal{F}) \geq c(\mathcal{E})c(\mathcal{F})$ for any pair $(\mathcal{E}, \mathcal{F})$ of dichotomies.

$$\begin{aligned} \lim_{n \rightarrow \infty} \sqrt[n]{b(\lambda | \mathcal{E}^n)} &= \lim_{n \rightarrow \infty} \sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)/2} = \\ &= \sup_n \sqrt[n]{b(\lambda | \mathcal{E}^n)} = \sup_n \sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)/2} = c(\mathcal{E}) . \end{aligned}$$

Corollary 6.17.

$$\lim_{n \rightarrow \infty} \sqrt[n]{1 - \delta(\mathcal{M}_i, \mathcal{E}^n)} = c(\mathcal{E}) .$$

Proof :

$$\sqrt[n]{1 - \delta(\mathcal{M}_i, \mathcal{E}^n)} = \sqrt[n]{2b(\frac{1}{2}, \frac{1}{2} | \mathcal{E}^n)} \rightarrow c(\mathcal{E}) .$$

□

If $c(\mathcal{E}) \neq c(\mathcal{F})$ then it follows from the inequalities :

$$\sqrt[n]{|\delta(\mathcal{E}^n, \mathcal{M}_a) - \delta(\mathcal{F}^n, \mathcal{M}_a)|} \leq \sqrt[n]{\Delta(\mathcal{E}^n, \mathcal{F}^n)} \leq \sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a) + \delta(\mathcal{F}^n, \mathcal{M}_a)}$$

that

$$\lim_{n \rightarrow \infty} \sqrt[n]{\Delta(\mathcal{E}^n, \mathcal{F}^n)} = c(\mathcal{E}) \vee c(\mathcal{F}) .$$

More generally, see [22] , this holds whenever $\mathcal{E} \nmid \mathcal{F}$.

The fact that $\delta(\mathcal{E}^n, \mathcal{M}_a)$ converges to zero with exponential speed if and only if $\mathcal{E} \nmid \mathcal{M}_i$ and $\mathcal{E} \nmid \mathcal{M}_a$ follows from :

Proposition 6.18.

$$\gamma(\mathcal{E})^2 / \sqrt[4]{4} \leq \sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)/2} \leq \inf_t H(t | \mathcal{E}) .$$

Proof :

The right " \leq " follows by the remark after theorem 6.3 while the left " \leq " follows by the remark after corollary 6.11. □

Corollary 6.19.

$$\gamma(\mathcal{C})^2 \leq c(\mathcal{C}) \leq \inf_t H(t|\mathcal{C})$$

Proof :

This follows from theorem 6.16 and proposition 6.18. □

It was shown by Chernoff, [6], that :

Theorem 6.20.*

For any dichotomy \mathcal{C} and any non degenerate prior distribution λ :

$$\lim_{n \rightarrow \infty} \frac{n}{\sqrt{b(\lambda|\mathcal{C}^n)}} = \inf_t H(t|\mathcal{C}) .$$

It follows then, by theorem 6.16, that the right " \leq " in corollary 6.18 may be replaced by " $=$ " i.e. :

Theorem 6.21.

$$c(\mathcal{C}) = \inf_t H(t|\mathcal{C}) .$$

The following property of the functional $\mathcal{C} \rightsquigarrow c(\mathcal{C})$ is immediate from this formula.

*) A proof of this theorem follows directly from the first part of the proof of theorem 7.2.

Corollary 6.22.

If \mathcal{C} and \mathcal{F} are dichotomies then $c(\mathcal{C} \times \mathcal{F}) \geq c(\mathcal{C}) \cdot c(\mathcal{F})$
and $c(\mathcal{C}^n) = c(\mathcal{C})^n$; $n=1,2,\dots$

Remark :

This follows, see footnote on page 6.13, also directly from theorem 6.16 and corollary 6.3.

The next two corollaries follow from remark 1 after theorem 6.3.

Corollary 6.23.

$$b(\lambda | \mathcal{C}) \leq (\lambda_1 \vee \lambda_2) c(\mathcal{C}) .$$

Corollary 6.24.

$$\delta(\mathcal{C}, \mathcal{M}_a)/2 \leq c(\mathcal{C}) .$$

Remark :

This inequality and the inequality $b(\lambda | \mathcal{C}) \leq c(\mathcal{C})$ follows immediately from theorem 6.16.

We shall now consider a few extensions of this result.

Let us first consider the asymptotic behaviour of minimum Baye's risk in other decision problems. It is known, see for example [7] or [19], that the minimum Baye's risk may often be expressed as functionals $\psi(\mathcal{C}) = \int \psi(dP_1, dP_2)$ where the function ψ is super linear* on R^2 . The function ψ is determined by the loss function.

It follows, since the standard measure of \mathcal{C}^n converges weakly to the standard measure of \mathcal{M}_a when $\mathcal{C} \rightarrow \mathcal{M}_i$, that

* i.e. $\psi(x+y) \geq \psi(x) + \psi(y)$ and $\psi(tx) = t\psi(x)$ when $t \geq 0$.

$\psi(\mathcal{E}^n) \rightarrow \psi(\mathcal{M}_a)$ as $n \rightarrow \infty$ provided $\mathcal{E} \vdash \mathcal{M}_i$. By theorem 2 in [19]:

$$0 \leq \psi(\mathcal{E}) - \psi(\mathcal{M}_a) \leq \frac{\delta(\mathcal{E}, \mathcal{M}_a)}{2} [\psi(1,0) + \psi(-1,0) + \psi(0,1) + \psi(0,-1)]$$

It follows, by replacing \mathcal{E} with \mathcal{E}^n and applying theorem 6.16, that

$$\limsup_n \sqrt[n]{\psi(\mathcal{E}^n) - \psi(\mathcal{M}_a)} \leq c(\mathcal{E}).$$

Suppose ψ is not affine on $[0, \infty[^2$. Put $\varphi(x) = \psi(1-x, x)$, $x \in [0, 1]$. Then φ is concave on $[0, 1]$ and for some $x_0 \in]0, 1[$: $\varphi(x_0) > (1-x_0)\varphi(0) + x_0\varphi(1)$. Let χ be the function on $[0, 1]$ which is linear on the intervals $[0, x_0]$ and $[x_0, 1]$ and which satisfies: $\chi(0) = \varphi(0)$, $\chi(x_0) = \varphi(x_0)$, $\chi(1) = \varphi(1)$. Then $\psi(\mathcal{E}) - \psi(\mathcal{M}_a) = \tilde{\psi}(\mathcal{E}) - \tilde{\psi}(\mathcal{M}_a)$ where $\tilde{\psi}(x) = \psi(x) - x_1\psi(1,0) - x_2\psi(0,1)$. Thus we may as well assume that $\varphi(0) = \varphi(1) = 0$ and then $\psi(\mathcal{E}) - \psi(\mathcal{M}_a) = \psi(\mathcal{E}) \geq \int \chi \left(\frac{dP_2}{dP_1 + P_2} \right) d(P_1 + P_2) = k b(\lambda, \mathcal{E})$ where $k = \varphi(x_0) \left(\frac{1}{x_0} + \frac{1}{1-x_0} \right)$ and $\lambda = \left(\frac{1}{1-x_0} \right) \left(\frac{1}{x_0} + \frac{1}{1-x_0} \right)^{-1}$.

Using theorem 6.16 once more we find altogether:

Theorem 6.25.

If ψ is super linear on \mathbb{R}^2 and not affine on $[0, \infty[^2$ then:

$$\lim_{n \rightarrow \infty} \sqrt[n]{\psi(\mathcal{E}^n) - \psi(\mathcal{M}_a)} \rightarrow c(\mathcal{E}).$$

Remark:

If $\psi(x) = \bigwedge_t \sum_{\theta=1}^2 L_{\theta}(t)x_{\theta}$, where L is the loss function, then

$\psi(\mathcal{E})$ is the minimum Baye's risk for the loss function L . The exceptional case is the situation where for some t_0 ; $L_\theta(t_0) \leq L_\theta(t)$ for $\theta = 1, 2$ and all t . In that case no observations are needed and the decision rule $x \rightsquigarrow t_0$ is "uniformly" optimal.

We shall for the remaining part of this section assume, unless otherwise stated, that :

$$(i) \quad \int_{P_1} \left(\frac{dP_2}{dP_1} \right) \text{ is non lattice}$$

$$(ii) \quad \inf_t H(t|\mathcal{E}) \text{ is obtained for } t = t^0 \text{ where } t^0 \text{ is non degenerate.}$$

In their paper [9], Effron and Truax obtained more accurate results on the asymptotic behaviour of $b(\lambda|\mathcal{E}^n)$. Using Edgeworth expansion they found, under assumptions (i) and (ii), that :

$$b(\lambda|\mathcal{E}^n) = \frac{\lambda_1^{t_1^0} \lambda_2^{t_2^0}}{t_1^0 t_2^0} \frac{1}{\sqrt{2\pi} \tau^{2r}} \frac{1}{\sqrt{n}} C(\mathcal{E})^n (1+o(1))$$

as $n \rightarrow \infty$. Here $\tau = C(\mathcal{E})^{-1} \left[\frac{d^2}{dt_2^2} H(1-t_2, t_2|\mathcal{E}) \right]_{t_2=t_2^0}$.

An exposition of this result, and related results, which is fitted to this framework may be found in [22].

We shall now use this result to get more information on the asymptotic behaviour of $\delta(\mathcal{E}^n, \mathcal{M}_a)$ and on the least favorable distribution for \mathcal{E}^n .

Put $A_n = C(\mathcal{E})^{-n} \sqrt{2\pi\tau^2} \sqrt{n}$. Effron's and Truax's expansion may then be written :

$$\lim_{n \rightarrow \infty} A_n b(1-\lambda, \lambda | \mathcal{E}^n) = \frac{(1-\lambda)^{1-t_0} \lambda^{t_0}}{(1-t_0) t_0}, \text{ where } \lambda \text{ is replaced by}$$

$(1-\lambda, \lambda)$ and $t_0 = t_2^0$. The functions $\lambda \rightsquigarrow b(\lambda | \mathcal{E}^n)$ and the function $\lambda \rightsquigarrow (1-\lambda)^{1-t_0} \lambda^{t_0}$ are all concave on $[0,1]$. It follows that the convergence is uniform in λ . Maximizing the left hand side and the right hand side w.r.t. λ we find that :

$$\lim_{n \rightarrow \infty} A_n \delta(\mathcal{E}^n, \mathcal{M}_a)/2 = \frac{(1-t_0)^{1-t_0} t_0^{t_0}}{(1-t_0) t_0}. \text{ Hence } \delta(\mathcal{E}^n, \mathcal{M}_a)/2 =$$

$= b(t^0 | \mathcal{E}^n)(1+o(1))$. Let $\mu^{(n)}$ be least favorable in \mathcal{E}^n ; i.e. $b(\mu^{(n)} | \mathcal{E}^n) = \delta(\mathcal{E}^n, \mathcal{M}_a)/2$. Then

$$A_n b(\mu^{(n)} | \mathcal{E}^n) - \frac{[\mu_1^{(n)}]^{t_1^0} [\mu_2^{(n)}]^{t_2^0}}{t_1^0 t_2^0} \rightarrow 0. \text{ Hence, by the asymptotic}$$

expression for $\delta(\mathcal{E}^n, \mathcal{M}_a)$:

$$\lim_{n \rightarrow \infty} [\mu_1^{(n)}]^{t_1^0} [\mu_2^{(n)}]^{t_2^0} \rightarrow [t_1^0]^{t_1^0} [t_2^0]^{t_2^0}$$

so that

$\mu^{(n)} \rightarrow t^0$. By a slight extension of this argument we find that $\delta(\mathcal{E}^n, \mathcal{M}_a)/2 = b(\mu^{(n)} | \mathcal{E}^n)(1+o(1))$ as $n \rightarrow \infty$, if and only if $\mu^{(n)} \rightarrow t^0$. This proves :

Theorem 6.26.

$$(i) \quad \delta(\mathcal{E}^n, \mathcal{M}_a)/2 = \frac{[t_1^0]^{t_1^0} [t_2^0]^{t_2^0}}{t_1^0 t_2^0} \frac{1}{\sqrt{2\pi\tau^2}} \frac{1}{\sqrt{n}} (1+o(1))$$

as $n \rightarrow \infty$.

(ii) t^0 is asymptotically least favorable in the sense that :

$$\delta(\mathcal{E}^n, \mathcal{M}_a)/2 = b(t^0 | \mathcal{E}^n)(1+o(1)) \quad \text{as } n \rightarrow \infty.$$

More generally :

$$(iii) \quad \delta(\mathcal{E}^n, \mathcal{M}_a)/2 = b(\mu^{(n)} | \mathcal{E}^n)(1+o(1)) \quad \text{as } n \rightarrow \infty$$

if and only if $\lim_n \mu^{(n)} = t^0$.

Remark :

The prior t^0 which minimizes $t \rightsquigarrow H(t | \mathcal{E})$ is, by this theorem, asymptotically least favorable.

Note next that, by proposition 6.2 :

$$\lim_{n \rightarrow \infty} \int [(1-\lambda) \wedge \lambda x] A_n K_n(dx) = \int [(1-\lambda) \wedge \lambda x] U_{t_2^0}(dx)$$

where $K_n = \int_{P_1^n} (dP_2^n/dP_1^n)$. It follows that

$$\lim_{n \rightarrow \infty} \int \varphi(x) A_n K_n(dx) = \int \varphi(x) U_{t_2^0}(dx) \quad \text{for any function}$$

φ on $[0, \infty[$ which is a linear combination of functions $x \rightsquigarrow (1-\lambda) \wedge \lambda x$; $\lambda \in [0, 1]$. It is not difficult to see that a function φ is a linear combination of functions $x \rightsquigarrow (1-\lambda) \wedge \lambda x$; $\lambda \in [0, 1]$, if and only if φ is polygonal, $\varphi(0) = 0$ and $\varphi(x) = \lim_{x \rightarrow \infty} \varphi(x)$ when x is sufficiently large. Hence, by the theory of weak convergence of measures :

Theorem 6.27.

$$\int \varphi(dP_2^n/dP_1^n) dP_1^n = \left[\int \varphi(x) U_{t_2^0}(dx) \right] \frac{1}{\sqrt{2\pi\tau^2}} \frac{1}{\sqrt{n}} O(\mathcal{E})^n(1+o(1))$$

as $n \rightarrow \infty$ for any bounded function φ on $[0, \infty[$ which is continuous a.e. Lebesgue and such that $\sup_{x>0} |\varphi(x)/x| < \infty$.

Or equivalently :

$$\int \rho\left(\frac{dP_2^n}{dP_1^n + dP_2^n}\right) d(P_1^n + P_2^n) = \int_0^1 \rho(x) \frac{(1-x)t_2^0}{x^{1+t_2^0}} dx \frac{1}{\sqrt{2\pi\tau^2}} \frac{1}{\sqrt{n}} (1+o(1))$$

as $n \rightarrow \infty$ for any function ρ on $[0, 1]$ which is continuous a.e. Lebesgue and such that $\sup_{1>x>0} |\rho(x)/x| < \infty$ and $\sup_{0<x<1} |\rho(x)/(1-x)| < \infty$.

Remark.

If ψ is sub linear or super linear on R^2 then $\rho(x) = \psi(1-x, x)$; $x \in [0, 1]$ satisfies the requirements of the theorem. Thus, by specializing to functions

$x \rightsquigarrow \bigwedge_{t=1}^2 \sum_{\theta} L_{\theta}(t) x_{\theta}$, we find asymptotic expressions for minimum Baye's risk in various decision problems.

As an application let us work out the asymptotic consequence of behaving according to a "wrong" prior distribution $(1-\lambda, \lambda)$.

Denote the "true" prior distribution by $(1-\mu, \mu)$. We shall assume that $\lambda, \mu \in]0, 1[$. The optimal Baye's test, δ , for the prior $(1-\lambda, \lambda)$ consists in rejecting or accepting according to whether $dP_2/dP_1 > \frac{1-\lambda}{\lambda}$ or $dP_2/dP_1 < \frac{1-\lambda}{\lambda}$. Expected loss is then $(1-\mu)P_1(\delta) + \mu P_2(1-\delta)$.

Now :

$$A_n P_1^n(dP_2^n/dP_1^n > \frac{1-\lambda}{\lambda}) = \int_{\frac{1-\lambda}{\lambda}}^{\infty} A_n K_n(dx) \rightarrow U_{t_0} \left[\frac{1-\lambda}{\lambda}, \infty \right] = \frac{1}{t_0} \left(\frac{\lambda}{1-\lambda} \right)^{t_0}$$

and

$$A_n P_2^n(dP_2^n/dP_1^n < \frac{1-\lambda}{\lambda}) = A_n \int_0^{\frac{1-\lambda}{\lambda}} x K_n(dx) \rightarrow \int_0^{\frac{1-\lambda}{\lambda}} x U_{t_0}(dx) = \frac{1}{1-t_0} \left(\frac{1-\lambda}{\lambda} \right)^{1-t_0}.$$

Hence :

$$\begin{aligned} (1-\mu) P_1^n(dP_2^n/dP_1^n > \frac{1-\lambda}{\lambda}) + \mu P_2^n(dP_2^n/dP_1^n < \frac{1-\lambda}{\lambda}) &= \\ &= c(\mathcal{E})^n \frac{1}{\sqrt{2\pi n \tau^2}} \left[\frac{1-\mu}{t_0} \left(\frac{\lambda}{1-\lambda} \right)^{t_0} + \frac{\mu}{1-t_0} \left(\frac{1-\lambda}{\lambda} \right)^{1-t_0} \right]. \end{aligned}$$

This should be compared with

$$b(\mu | \mathcal{E}^n) \sim c(\mathcal{E})^n \frac{1}{\sqrt{2\pi n \tau^2}} \frac{(1-\mu)^{1-t_0} \mu^{t_0}}{(1-t_0) t_0}$$

which is the minimum Baye's risk we would have obtained if we had chosen the right prior distribution.

This proves :

Proposition 6.28.

Let, δ_n , $n = 1, 2, \dots$, be a test in \mathcal{E}^n which achieves minimum Baye's risk w.r.t. the nondegenerate prior distribution λ . Then the ratio of the Baye's risk of this test w.r.t. the nondegenerate prior distribution μ and the minimum Baye's risk for the prior μ converges as $n \rightarrow \infty$ to

$$\frac{t_1^0 t_2^0}{\mu_1^0 \mu_2^0} \left[\frac{\mu_1}{t_2^0} \left(\frac{\lambda_2}{\lambda_1} \right)^{t_2^0} + \frac{\mu_2}{t_1^0} \left(\frac{\lambda_1}{\lambda_2} \right)^{t_1^0} \right].$$

Remark 1.

Note that the exponential rate of convergence to zero is not affected by choosing the wrong prior distribution. The risk is still : $c(\mathcal{C})^n \{[1+o(1)]^n\}$. This follows quite generally, (assumptions (i) and (ii) are not needed), from the simple inequalities :

$$\begin{aligned} \left[\frac{1-\mu}{1-\lambda} \wedge \frac{\mu}{\lambda} \right] \left[(1-\lambda)P_1^n(\delta_n) + \lambda P_2^n(1-\delta_n) \right] &\leq (1-\mu)P_1^n(\delta_n) + \mu P_2^n(1-\delta_n) \leq \\ &\leq \left[\frac{1-\mu}{1-\lambda} \vee \frac{\mu}{\lambda} \right] \left[(1-\lambda)P_1^n(\delta_n) + \lambda P_2^n(1-\delta_n) \right] . \end{aligned}$$

Not only is the exponential rate preserved. The risk is still of the form :

$$c(\mathcal{C})^n \frac{1}{\sqrt{n}} \cdot \text{constant} .$$

Thus the asymptotic consequences of choosing the wrong prior appears only in the constant.

Remark 2.

The limiting value of the ratio obtains, as it should, its maximal value 1 when $\lambda = \mu$.

The exponential rate of convergence of the power of the most powerful level α test for testing " $\theta=1$ " against " $\theta=2$ " was determined by D.D. Joshi [1957, L'information en statistique mathématique et dans la théorie des communications. Thèse, Faculté des Sciences de l'Université de Paris, June.] An asymptotic expansion, based on Edgeworth expansions, for this power was given by B. Effron [1967, The power of the likelihood ratio test. AMS 38, 802-806]. See [22] for an exposition adapted to our framework. Problems concerning exponential rates of convergence for risk functions which are optimal w.r.t. side conditions of this type will not be discussed in this paper.

7. Replicated experiments when the parameter set is finite.

How fast does the content of information in n replicates of an experiment \mathcal{E} increase when $n \uparrow \infty$? We shall in this section investigate this problem when Θ is finite. It is, in view of the fact that pairwise sufficiency implies sufficiency, not too surprising that the problem may be reduced to the same problem for dichotomies.

We extend the definition of the constant $C(\mathcal{E})$ in section 6 by defining :

$$C(\mathcal{E}) = \max_{\theta_1 \neq \theta_2} \inf_{0 < t < 1} \int dP_{\theta_1}^{1-t} dP_{\theta_2}^t .$$

Thus :

$$C(\mathcal{E}) = \max_{\theta_1 \neq \theta_2} C(\mathcal{E}_{\{\theta_1, \theta_2\}}) .$$

The parameter set Θ will, unless otherwise stated, be assumed finite throughout this section.

Consider now an experiment $\mathcal{E} = (P_\theta; \theta \in \Theta)$.

Let ψ be sub linear on R^Θ and let F be a non empty sub set of Θ . Then, by sub linearity :

$$\psi(z) = \psi(\sum_{\theta} z_{\theta} e^{\theta}) \leq \psi(\sum_{F} z_{\theta} e^{\theta}) + \sum_{F^c} z_{\theta} \psi(e^{\theta}) .$$

Let S denote the standard measure of \mathcal{E} . Then :

$$\begin{aligned} \psi(\mathcal{M}_a) - \psi(\mathcal{E}) &= \sum_{\theta} \psi(e^{\theta}) - \int \psi dS = \int [\sum_{\theta} z_{\theta} \psi(e^{\theta}) - \psi(\sum_{\theta} z_{\theta} e^{\theta})] S(dz) \\ &\geq \int [\sum_{F} z_{\theta} \psi(e^{\theta}) - \psi(\sum_{F} z_{\theta} e^{\theta})] S(dz) = \psi_F(\mathcal{M}_a) - \psi_F(\mathcal{E}) \end{aligned}$$

where $\psi_F(z) = \psi(\sum_F z_\theta e^\theta)$. If, in particular, $F = \{\theta_1, \theta_2\}$ then :

$$\psi(\mathcal{M}_a) - \psi(\mathcal{G}) \geq \psi_{\{\theta_1, \theta_2\}}(\mathcal{M}_a) - \psi_{\{\theta_1, \theta_2\}}(\mathcal{G}).$$

Substituting \mathcal{G}^n for \mathcal{G} and applying theorem 6.25 we find, provided $\psi_{\{\theta_1, \theta_2\}}$ is not affine on $[0, \infty[^\Theta$, that

$$\liminf_n \sqrt[n]{\psi(\mathcal{M}_a) - \psi(\mathcal{G}^n)} \geq c(\mathcal{G}_{\{\theta_1, \theta_2\}}).$$

Suppose now that this provision is satisfied for all two points sets $\{\theta_1, \theta_2\}$. Then :

$$\liminf_n \sqrt[n]{\psi(\mathcal{M}_a) - \psi(\mathcal{G}^n)} \geq c(\mathcal{G}).$$

The provision above is obviously satisfied for any function ; $z \rightsquigarrow \sum_\theta \lambda_\theta z_\theta$ where λ is a prior distribution on Θ such that $\lambda_\theta > 0$ for all θ . Suppose λ satisfies this condition. Then the above result imply that :

$$\liminf_n \sqrt[n]{b(\lambda | \mathcal{G}^n)} \geq c(\mathcal{G}).$$

Hence, by proposition 3.8 and theorem 6.16 :

$$\begin{aligned} \limsup_n \sqrt[n]{b(\lambda | \mathcal{G}^n)} &\leq \\ &\leq \limsup_{n \rightarrow \infty} \sqrt{\frac{1}{2} \sum_{\theta_1 \neq \theta_2} (\lambda_{\theta_1} + \lambda_{\theta_2}) b\left(\frac{\lambda_{\theta_1}}{\lambda_{\theta_1} + \lambda_{\theta_2}}, \frac{\lambda_{\theta_2}}{\lambda_{\theta_1} + \lambda_{\theta_2}} | \mathcal{G}_{\{\theta_1, \theta_2\}}^n\right)} \\ &\leq \max_{\theta_1 \neq \theta_2} c(\mathcal{G}_{\{\theta_1, \theta_2\}}) = c(\mathcal{G}). \end{aligned}$$

It follows that $\sqrt[n]{b(\lambda | \mathcal{E}^n)} \rightarrow c(\mathcal{E})$ as $n \rightarrow \infty$. Hence, by corollary 3.2 and proposition 3.7 :

$$\begin{aligned} c(\mathcal{E}) &= \lim_n \sqrt[n]{b(\text{uniform} | \mathcal{E}^n)} \leq \liminf_n \sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)/2} \\ &\leq \limsup_n \sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)/2} \leq \limsup_n \sqrt[n]{[\#\Theta] b(\text{uniform} | \mathcal{E}^n)} = \\ &= c(\mathcal{E}) , \end{aligned}$$

so that $\lim_n \sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)} = c(\mathcal{E})$.

By the sub linear function criterion :

$$\begin{aligned} \limsup_n \sqrt[n]{\psi(\mathcal{M}_a) - \psi(\mathcal{E}^n)} &\leq \limsup_n \sqrt[n]{\sum_{\theta} \frac{1}{2} [\psi(e^{\theta}) + \psi(e^{-\theta})] \delta(\mathcal{E}^n, \mathcal{M}_a)} \\ &\leq \limsup_n \sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)} = c(\mathcal{E}) \quad \text{for any sub linear function } \psi \\ &\text{on } R^{\Theta} . \end{aligned}$$

Finally, by corollary 5.5 :

$$\lim_n \sqrt[n]{2 - \frac{2}{m} - \delta(\mathcal{M}_1, \mathcal{E}^n)} = c(\mathcal{E}) , \text{ where } m = \#\Theta .$$

Altogether we have proved :

Theorem 7.1.

Let $\mathcal{E} = (P_{\theta}; \theta \in \Theta)$ be an experiment with finite parameter set.
Then :

- (i) $\lim_{n \rightarrow \infty} \sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)} = c(\mathcal{E}) .$
- (ii) $\lim_{n \rightarrow \infty} \sqrt[n]{2 - \frac{2}{m} - \delta(\mathcal{M}_1, \mathcal{E}^n)} = c(\mathcal{E}) \text{ where } m = \#\Theta .$
- (iii) $\lim_{n \rightarrow \infty} \sqrt[n]{b(\lambda | \mathcal{E}^n)} = c(\mathcal{E})$ provided $\lambda_{\theta} > 0$ for all θ .

- (iv) $\limsup_{n \rightarrow \infty} \sqrt[n]{\psi(\mathcal{M}_a) - \psi(\mathcal{G}^n)} \leq c(\mathcal{G})$ for any sub linear function ψ
- (v) $\lim_{n \rightarrow \infty} \sqrt[n]{\psi(\mathcal{M}_a) - \psi(\mathcal{G}^n)} = c(\mathcal{G})$ for any sub linear function ψ on R^{Θ} such that ψ is non of the maps $z \mapsto \psi(z_{\theta_1} e^{\theta_1} + z_{\theta_2} e^{\theta_2})$; $\theta_1 \neq \theta_2$ are affine on $[0, \infty[^{\Theta}$.
-

Remark:

If $\psi(x) = \bigvee_t \sum_{\theta} \lambda_{\theta} U_{\theta}(t) x_{\theta}$ where T is a decision space and U is the utility function, then (iv) describes the exponential rate of convergence to $\sum_{\theta} \lambda_{\theta} \bigvee_t U_{\theta}(t)$ of maximum Baye's utility. The exceptional case is precisely the situation where for some two point set $\{\theta_1, \theta_2\}$, no observations are needed when it is known that $\theta \in \{\theta_1, \theta_2\}$.

Although theorem 7.1 yields the exact rate of exponential convergence in many situations, there are situations of interest where the condition in (v) is not satisfied. Consider, for example, the problem of catching θ with an r -point confidence set. Then the minimax probability of not covering the true value is:

$$\kappa_r(\mathcal{G}) = 1 - \inf_{\lambda} \left\| \bigvee_U \sum_{\theta \in U} \lambda_{\theta} P_{\theta} \right\|$$

- *) Thus (v) is not applicable to expressions like $\left\| \bigwedge_{\theta} \lambda_{\theta} P_{\theta}^n \right\| = \left\| \bigvee_{\theta} (-\lambda_{\theta}) P_{\theta}^n \right\|$ and, in fact, $\left\| \bigwedge_{\theta} \lambda_{\theta} P_{\theta}^n \right\| \leq \prod_{\theta} \lambda_{\theta}^{t_{\theta}} H_{\mathcal{G}}(t)^n$ for any pair (λ, t) of prior distributions on Θ .

where U runs through all r -point subsets of Θ . Then $\kappa_1 = \delta(\mathcal{E}, \mathcal{M}_a)/2$ and it is easily seen that κ_r is monotonically decreasing in r .

Let us briefly consider the asymptotic behaviour of these quantities when Θ is finite:

Theorem 7.2

Suppose Θ is finite and put $m = \#\Theta$. Define for each experiment \mathcal{E} and each integer $r \in \{1, 2, \dots, m-1\}$ the quantity $\kappa_r(\mathcal{E})$ as above. Then:

$$\lim_{n \rightarrow \infty} \sqrt[n]{\kappa_r(\mathcal{E}^n)} = \max_W \inf_{t \in \Lambda_W} \int_W \prod (dP_\theta)^{t_\theta}$$

where

- (i) W runs through all $(r+1)$ -point subsets of Θ
- and
- (ii) Λ_W , for each W is the set of all prior distributions on Θ which are supported by W .

Furthermore: The n -th root of the minimum Bayes's probability of not covering the true value of θ .

$$\sqrt[n]{1 - \left\| \bigvee_{U} \sum_{\theta \in U} \lambda_\theta P_\theta^n \right\|}$$

converges to the same limit, provided $\lambda_\theta > 0$ for all $\theta \in \Theta$.

Remark :

Putting $m = 2$ and $r = 1$ we see that the last statement generalizes Chernoff's result, theorem 6.20.

Proof of theorem 7.2 : Let us write $\Theta = \{1, 2, \dots, m\}$. The proof is completed in two steps.

First step :

Claim 1 : $\sqrt[n]{\|\bigwedge_{\theta} a_{\theta}^n P_{\theta}^n\|} \rightarrow \inf_{t \in \Lambda} \prod_{\theta} a_{\theta}^t H(t)$ when $a_{\theta} \geq 0$ for all θ

and $H(t) = H(t|\mathcal{G})$.

Proof of claim 1 :

It suffices, in view of the inequalities :

$$\|\bigwedge_{\theta} a_{\theta}^n P_{\theta}^n\| \leq \prod_{\theta} a_{\theta}^{nt} H(t)^n ; t \in \Lambda$$

to show that :

$$(\S) \quad \liminf_n \sqrt[n]{\|\bigwedge_{\theta} a_{\theta}^n P_{\theta}^n\|} \geq \inf_t \prod_{\theta} a_{\theta}^t H(t) \text{ when } a_{\theta} > 0 \text{ for all } \theta ,$$

$a_m = 1$ and $\inf_t H(t) > 0$. Let S denote the standard measure of $\mathcal{G} = (P_{\theta} : \theta \in \Theta)$. Then $S(\Lambda) > 0$. Suppose we have proved (\S) when $\text{supp } S$ is bounded away from the boundary of Λ . Then (\S) follows by :

- 1) truncating S to the subset of Λ consisting of all points x whose distance to the boundary is at least $\epsilon > 0$

and

- 2) applying, as $\epsilon \rightarrow 0$, the minimax theorem in Chernoff's paper [6]. [See also page P.D. 2.6 in [22]].

Let us now assume that $\text{supp } S$ is bounded away from the boundary. Then H is analytic on Λ . Let $\mathcal{F} = (Q_1, Q_2, \dots, Q_m)$ be a homogenous experiment such that $\mathcal{L}_{Q_m}(\log \frac{dQ_i}{dQ_m} ; i=1, 2, \dots, m-1)$ is absolutely continuous and such that $H(\cdot|\mathcal{F})$ is analytic.

Then $\mathcal{G} \times \mathcal{F}$ have the same properties. Suppose we have proved claim 1 for all experiments having these properties.

Then :

$$\begin{aligned} & \liminf_n \sqrt[n]{\|\bigwedge_i a_i^n P_i^n\|} \\ & \geq \liminf_n \sqrt[n]{\|\bigwedge_i a_i^n [P_i \times Q_i]^n\|} = \inf_t [\prod_{\theta} a_{\theta}^t H(t) H(t|\mathcal{F})] \\ & \geq [\inf_t H(t|\mathcal{F})] [\inf_t \prod_{\theta} a_{\theta}^t H(t)] . \end{aligned}$$

Replacing \mathcal{F} by $\frac{1}{N}\mathcal{F} + (1-\frac{1}{N})\mathcal{A}_i$ in the last expression and letting $N \rightarrow \infty$ we get :

$$\liminf_n \sqrt[n]{\|\bigwedge_i a_i^n P_i^n\|} \geq \inf_t a^t H(t) \text{ where } a^t = \prod_{\theta} a_{\theta}^t .$$

It follows that we may assume that H is analytic on all of Λ and that

$$\mathcal{L}_{P_m} \left(\log \frac{dP_i}{dP_m} ; i=1,2,\dots,m-1 \right)$$

is absolutely continuous on R^{m-1} . Put

$$F = \mathcal{L}_{P_m} \left(\log \frac{dP_i}{dP_m} ; i=1,2,\dots,m-1 \right) ,$$

and

$$H_0 = \inf_t a^t H(t) .$$

Then

$$H(t) = \int e^{\langle t, z \rangle} F(dz)$$

where

$$\langle t, z \rangle = \sum_{i=1}^{m-1} t_i z_i ,$$

and

$$\|\wedge_i a_i^n P_i^n\| = \int \left[\wedge_{i=1}^{m-1} e^{z_i - n\tilde{a}_i} \wedge 1 \right] F(dz)$$

where $\tilde{a}_i = -\log a_i$; $i=1,2,\dots,m$.

Let t^0 be a point in Λ where $a^t H(t)$ achieves minimum i.e. $H_0 = H(t^0)$. We may, without loss of generality, assume that $t_m^0, t_{m-1}^0, \dots, t_{m-k+1}^0 > 0$ and $t_{m-k}^0 = \dots = t_1^0 = 0$ where $k \geq 2$.

Differentiating we find :

$$\int (z_i - \tilde{a}_i) e^{\langle t^0, z - \tilde{a} \rangle} F(dz) = 0 \quad \text{or} \quad \geq 0$$

as $i \geq m-k+1$ or $i \leq m-k$.

$$\text{Put } \zeta_i = \frac{1}{H_0} \int z_i e^{\langle t^0, z - \tilde{a} \rangle} F(dz) \quad ; \quad i=1,2,\dots,m-1$$

and

$$\sigma_{ij} = \frac{1}{H_0} \int (z_i - \zeta_i)(z_j - \zeta_j) e^{\langle t^0, z - \tilde{a} \rangle} F(dz) \quad ; \quad i,j=1,2,\dots,m-1.$$

Then $\zeta_i = 0$ or ≥ 0 as $i \geq m-k+1$ or $i < m-k+1$.

Let ϕ denote the multivariate normal distribution with expectation vector ζ and covariance matrix σ .

Introduce a new measure G by :

$$\frac{dG}{dF} = H_0^{-1} e^{\langle t^0, z - \tilde{a} \rangle}.$$

$$\text{Then :} \quad \int z_i G(dz) = \zeta_i$$

$$\text{and} \quad \int (z_i - \zeta_i)(z_j - \zeta_j) G(dz) = \sigma_{ij}.$$

Let \tilde{K}_n be the distribution of $\sum_{j=1}^n (Z^j - \zeta) / \sqrt{n}$ when Z^1, Z^2, \dots, Z^n are independently and identically distributed, each distributed according to G .

Then :

$$\int z_i K_n(dz) = 0 \quad ; \quad i=1, 2, \dots, m-1$$

and

$$\int z_i z_j K_n(dz) = \sigma_{ij} \quad ; \quad i, j = 1, 2, \dots, m-1.$$

By the central limit theorem :

$$\int h d K_n \rightarrow \int h d \Phi$$

when h is continuous and bounded on R^{m-1} .

We find easily :

$$\begin{aligned} \left\| \bigwedge_{i=1}^m a_i^n P_i^n \right\| &= \int \left[\bigwedge_{i=1}^{m-1} e^{z_i - \tilde{a}_i} \wedge 1 \right] F^{n*}(dz) = \\ &= H_0^n \int \left[\bigwedge_{i=1}^{m-1} e^{z_i - \tilde{a}_i} \wedge 1 \right] e^{-\langle t^0, z - \tilde{a} \rangle} G^{n*}(dz) = \\ &= H_0^n \int \left[\bigwedge_{i=1}^{m-1} e^{\sqrt{n} z_i + n(\zeta_i - \tilde{a}_i)} \wedge 1 \right] e^{-\sqrt{n} \langle t^0, z \rangle} K_n(dz) \geq \\ &\geq H_0^n K_n(A) \int e^{-\sqrt{n} \langle t^0, z \rangle} \tilde{K}_n(dz) \end{aligned}$$

where

$$A = \{z : z_1 > 0, z_2 > 0, \dots, z_{m-1} > 0\}$$

and

\tilde{K}_n is the conditional distribution of Z given that $\mathcal{L}(Z) = K_n$ and that $Z \in A$. $K_n(A)$ is, by the central limit theorem, positive when n is sufficiently large and then \tilde{K}_n is well defined. Assume for the remaining part of the proof of claim 1 that n is restricted to $\{n : K_n(A) > 0\}$.

[By weak convergence : $K_n(A) \rightarrow \Phi(A) > 0$.]

By Jensen's inequality :

$$\int e^{-\sqrt{n}\langle t^0, \zeta \rangle} \tilde{K}_n(dz) \geq e^{-\sqrt{n}\langle t^0, \int z \tilde{K}_n(dz) \rangle} .$$

Now :

$$\left| \int_A z_i K_n(dz) \right| \leq \int |z_i| K_n(dz) \leq \sqrt{\sigma_{i,i}} ; i > m-k .$$

It follows that :

$$\sup_n \max_{i > m-k} \left| \int z_i \tilde{K}_n(dz) \right| < \infty .$$

Hence :

$$\liminf_n \sqrt[n]{\left\| \bigwedge_{i=1}^m a_i P_i^n \right\|} \geq H_0 \liminf_n \left[\frac{n}{\sqrt{K_n(A)}} e^{-n^{-\frac{1}{2}} \langle t^0, \int z \tilde{K}_n(dz) \rangle} \right] = H_0 .$$

This completes the proof of our first claim.

It follows in particular that $\sqrt[n]{\left\| \bigwedge_{\theta} a_{\theta} P_{\theta}^n \right\|} \rightarrow \inf_t H(t)$ when $a_{\theta} > 0$

for all θ .

Second step :

Claim 2 :

$$\begin{aligned} & \binom{m-1}{r}^{-1} \sum_{i_1 < i_2 < \dots < i_{r+1}} \kappa_r(\mathcal{G}_{\{i_1, i_2, \dots, i_{r+1}\}}) \leq \kappa_r(\mathcal{G}) \leq \\ & \leq \sum_{i_1 < i_2 < \dots < i_{r+1}} \kappa_r(\mathcal{G}_{\{i_1, i_2, \dots, i_{r+1}\}}) . \end{aligned}$$

Proof of claim 2 :

We may write :

$$\kappa_r(\mathcal{E}) = \sup_{\lambda} \left\| \bigwedge_U \sum_{U^c} \lambda_{\theta} P_{\theta} \right\| .$$

Let $y_1 \leq y_2 \leq y_3 \leq \dots \leq y_m$.

Then

$$\bigwedge_U \sum_{U^c} y_{\theta} = y_1 + \dots + y_s \quad \text{where } s = m-r .$$

Claim 2 follows now immediately from the identity :

$$\sum \{y_{i_1} : i_1 < i_2 < \dots < i_{r+1}\} = \binom{m-1}{r} y_1 + \binom{m-2}{r} y_2 + \dots + \binom{r}{r} y_s .$$

The proof of the theorem is now completed by inserting \mathcal{E}^n for \mathcal{E} , taking n -th root, letting $n \rightarrow \infty$, and using claim 1 .

□

If the dichotomies $\mathcal{E}_{\{\theta_1, \theta_2\}}$; $\theta_1 \neq \theta_2$ all satisfies conditions (i) and (ii) in section 6, then it is possible to obtain more accurate results. The constant $\# \Theta$ will, however, appear in unpleasant ways and further study should be undertaken.

Let $\mathcal{E} = (P_{\theta}; \theta \in \Theta)$ and $\mathcal{F} = (Q_{\theta}; \theta \in \Theta)$ be two experiments. If

$\mathcal{E}_{\{\theta_1, \theta_2\}} \uparrow \mathcal{F}_{\{\theta_1, \theta_2\}}$ when $\theta_1 \neq \theta_2$, then by theorem 6.16,

and theorem 3 in [22] :

$$\begin{aligned}
\liminf_n \sqrt[n]{\Delta(\mathcal{E}^n, \mathcal{F}^n)} &\geq \liminf_n \sqrt[n]{\max_{\theta_1 \neq \theta_2} \Delta(\mathcal{E}_{\{\theta_1, \theta_2\}}^n, \mathcal{F}_{\{\theta_1, \theta_2\}}^n)} \\
&= \max_{\theta_1 \neq \theta_2} \max \{c(\mathcal{E}_{\{\theta_1, \theta_2\}}), c(\mathcal{F}_{\{\theta_1, \theta_2\}})\} \\
&= c(\mathcal{E}) \vee c(\mathcal{F}) .
\end{aligned}$$

In any case :

$$\begin{aligned}
\limsup_n \sqrt[n]{\Delta(\mathcal{E}^n, \mathcal{F}^n)} &\leq \limsup_n \sqrt[n]{\Delta(\mathcal{E}^n, \mathcal{M}_a) + \Delta(\mathcal{F}^n, \mathcal{M}_a)} \\
&= c(\mathcal{E}) \vee c(\mathcal{F}) .
\end{aligned}$$

Hence :

Proposition 7.3.

Let $\mathcal{E} = (P_\theta; \theta \in \Theta)$ and $\mathcal{F} = (Q_\theta; \theta \in \Theta)$

be two experiments.

Then

$$\limsup_n \sqrt[n]{\Delta(\mathcal{E}^n, \mathcal{F}^n)} \leq c(\mathcal{E}) \vee c(\mathcal{F}) .$$

If $\mathcal{E}_{\{\theta_1, \theta_2\}} \nmid \mathcal{F}_{\{\theta_1, \theta_2\}}$ when $\theta_1 \neq \theta_2$ then moreover :

$$\lim_{n \rightarrow \infty} \sqrt[n]{\Delta(\mathcal{E}^n, \mathcal{F}^n)} = c(\mathcal{E}) \vee c(\mathcal{F}) .$$

Remark 1.

It is our conjecture that $\lim_n \sqrt[n]{\Delta(\mathcal{E}^n, \mathcal{F}^n)} = c(\mathcal{E}) \vee c(\mathcal{F})$ whenever $\mathcal{E} \nmid \mathcal{F}$. In particular this should hold when \mathcal{E} and \mathcal{F} are pairwise equivalent but not equivalent.

Remark 2.

Another interesting, and open, problem is that of describing the asymptotic behavior of $\delta(\mathcal{E}^n, \mathcal{F}^n)$. A few relevant results are given on pages 3.6 and 3.7 in [22].

If the map $\theta \rightsquigarrow P_\theta$ is not 1-1 then $C(\mathcal{E}) = 1$ and $\delta(\mathcal{E}^n, \mathcal{M}_a) \geq 1$ for all n . The obvious way out is to replace the parameter set Θ by the set $\{P_\theta; \theta \in \Theta\}$.

Let, for each partition $\{\theta_1, \theta_2, \dots, \theta_r\}$ of Θ into non empty sets, $\mathcal{M}(\{\theta_1, \theta_2, \dots, \theta_r\})$ denote any experiment $(Q_\theta; \theta \in \Theta)$ such that $Q_{\theta_1} = Q_{\theta_2}$ or $Q_{\theta_1} \wedge Q_{\theta_2} = 0$ as θ_1 and θ_2 are equivalent or non equivalent according to the partition. Clearly any two experiments of this type are equivalent and we shall not bother to distinguish between them. Note that

$$\mathcal{M}_a = \mathcal{M}(\{\{\theta\}\}; \theta \in \Theta)$$

and

$$\mathcal{M}_i = \mathcal{M}(\{\Theta\}).$$

Call an experiment \mathcal{E} idempotent if $\mathcal{E}^2 \sim \mathcal{E}$. The idempotent experiments are characterized in :

Proposition 7.4.

The following conditions on an experiment \mathcal{E} are all equivalent :

- (i) \mathcal{E} is idempotent
- (ii) $\mathcal{E} \sim \lim_n \mathcal{F}^n$ for some experiment \mathcal{F} .

- (iii) $\mathcal{C} \sim \mathcal{M}(\{\theta_1, \theta_2, \dots, \theta_r\})$ for some partition $\{\theta_1, \theta_2, \dots, \theta_r\}$ of Θ into non empty sub sets $\theta_1, \theta_2, \dots, \theta_r$.
-

Proof :

- (i) \Rightarrow (ii) : If $\mathcal{C} \sim \mathcal{C}^2$, then by induction $\mathcal{C} \sim \mathcal{C}^n$;
 $n=1, 2, \dots$.

- (ii) \Rightarrow (iii): Write $\mathcal{F} = (Q_\theta; \theta \in \Theta)$. Then $\mathcal{F}^n \rightarrow \mathcal{M}(\{\theta_1, \theta_2, \dots, \theta_r\})$ where $\{\theta_1, \theta_2, \dots, \theta_r\}$ is defined by the equivalence relation :

$$\theta_1 \sim \theta_2 \stackrel{\text{def}}{\iff} Q_{\theta_1} = Q_{\theta_2}.$$

- (iii) \Rightarrow (i) : Put $\mathcal{C} = (P_\theta; \theta \in \Theta)$, and let \sim be the equivalence relation defined by the partition $(\theta_1, \dots, \theta_r)$. Then $P_{\theta_1} = P_{\theta_2}$ or $P_{\theta_1} \wedge P_{\theta_2} = 0$ as $\theta_1 \sim \theta_2$ or $\theta_1 \nmid \theta_2$. Hence $P_{\theta_1}^2 = P_{\theta_2}^2$ or $P_{\theta_1}^2 \wedge P_{\theta_2}^2 = 0$ as $\theta_1 \sim \theta_2$ or $\theta_1 \nmid \theta_2$. Thus $\mathcal{C}^2 \sim \mathcal{M}(\{\theta_1, \dots, \theta_r\}) \sim \mathcal{C}$.



Corollary 7.5.

Let $\mathcal{G} = (P_\theta; \theta \in \Theta)$ and let $\{\Theta_1, \Theta_2, \dots, \Theta_r\}$ be the partition of Θ into non empty sub sets which is induced by the equivalence relation : $\{(\theta_1, \theta_2) : P_{\theta_1} = P_{\theta_2}\}$.

Let $\theta_i \in \Theta_i$; $i=1,2,\dots,r$.

Then :

$$(i) \quad \lim_{n \rightarrow \infty} \sqrt[n]{\delta(\mathcal{G}^n, \mathcal{M}(\{\Theta_1, \Theta_2, \dots, \Theta_r\}))} = c(\mathcal{G}_{\{\theta_1, \theta_2, \dots, \theta_r\}})$$

and

$$(ii) \quad \lim_{n \rightarrow \infty} \sqrt{2 - \frac{2}{r} - \delta(\mathcal{M}_i, \mathcal{G}^n)} = c(\mathcal{G}_{\{\theta_1, \theta_2, \dots, \theta_r\}}) .$$

Proof :

This follows directly from theorem 7.1.

□

8. Replicated experiments when the parameter set is countable.

In order to generalize the results in section 7 to the case of countable parameter sets we shall have to search for sequences $\{\hat{\theta}_n\}$ estimators of θ such that $\hat{\theta}_n$ is based on \mathcal{G}^n and is such that $P_{\theta}^n(\hat{\theta}_n \neq \theta)$ is small. The definition of the crucial quantity $C(\mathcal{G})$ in section 7 is extended to the countable case by defining:

$$C(\mathcal{G}) = \sup_{\theta_1 \neq \theta_2} \inf_{0 < t < 1} \int dP_{\theta_1}^{1-t} dP_{\theta_2}^t$$

where $\mathcal{G} = (P_{\theta}; \theta \in \Theta)$. Thus $C(\mathcal{G}) = \sup_{\theta_1 \neq \theta_2} C(\mathcal{G}_{\{\theta_1, \theta_2\}})$, and

$C(\mathcal{G}^n) = C(\mathcal{G})^n$, $n = 1, 2, \dots$. Put also $\tau(\mathcal{G}) = \inf_n \sqrt[n]{\delta(\mathcal{G}^n, \mathcal{M}_a)/2}$ and $\sigma(\mathcal{G}) = C(\mathcal{G}) \vee \tau(\mathcal{G})$. The possibilities of exponential convergence are completely characterized by $\sigma(\mathcal{G})$. This is a consequence of:

Theorem 8.1.

Let $\mathcal{G} = (P_{\theta}; \theta \in \Theta)$ be an experiment with countable parameter set Θ and let X_1, X_2, \dots be independent observations of \mathcal{G} .

Then there is a sequence $(\hat{\theta}_n(X_1, X_2, \dots, X_n); n = 1, 2, \dots)$ of estimators of θ such that

$$\lim_{n \rightarrow \infty} \sqrt[n]{\sup_{\theta} P_{\theta}^n(\hat{\theta}_n \neq \theta)} = \sigma(\mathcal{G}).$$

On the other hand, for any sequence $\{\hat{\theta}_n(X_1, X_2, \dots, X_n); n = 1, 2, \dots\}$ of estimators of θ do we have:

$$\liminf_{n \rightarrow \infty} \sqrt[n]{\sup_{\theta} P_{\theta}^n(\hat{\theta}_n \neq \theta)} \geq \sigma(\mathcal{G}).$$

Furthermore:

$$\lim_{n \rightarrow \infty} \sqrt[n]{\delta(\mathcal{G}^n, \mathcal{M}_a)} = \sigma(\mathcal{G}) .$$

Remark:

If Θ is finite then by theorem 6.16 $\tau(\mathcal{G}) \leq C(\mathcal{G})$ so that $\sigma(\mathcal{G}) = C(\mathcal{G})$. Note also that $C(\mathcal{G}) = \tau(\mathcal{G}) = \sigma(\mathcal{G}) = 1$ when the map $\theta \rightarrow P_\theta$ is not 1-1.

Proof :

Let μ be any measure dominating $(P_\theta : \theta \in \Theta)$ and put, for each θ , $f_{\theta,n} = dP_\theta^n/d\mu^n$; $n = 1, 2, \dots$. Let us first assume that $\delta(\mathcal{G}, \mathcal{M}_a) < 2$. There is, by assumption, a randomization M from \mathcal{G} to Θ so that $P_\theta M(\theta) \geq \alpha > 0$ where $\alpha = 1 - \delta(\mathcal{G}, \mathcal{M}_a)/2$. Put $Q_\theta = P_\theta M$ and $\mathcal{F} = (Q_\theta : \theta \in \Theta)$. Let $X_1, X_2, \dots, X_n, \dots$ be independent observations of \mathcal{G} and let Y_1, Y_2, \dots be independent Θ valued variables such that

$$\mathcal{L}(Y_i | X_i) = M(\cdot | X_i)$$

Thus Y_1, Y_2, \dots are independent observations of \mathcal{F} . Put for each $\xi \in]0, 1[$:

$$F_\theta(\xi) = \{\theta' : Q_\theta(\theta') \geq \xi\}$$

and

$$A_n(\xi) = \{\theta' : h_n(\theta') \geq \xi\}$$

where, for each subset F of Θ , $h_n(F) = \frac{1}{n} \# \{i : 1 \leq i \leq n, y_i \in F\}$. Note that the set $A_n(\xi)$ is a random subset of Θ based on (X_1, X_2, \dots, X_n) . The sets $F_\theta(\xi)$ and $A_n(\xi)$ are both necessarily finite and $\#F_\theta(\xi) \leq \lceil \frac{1}{\xi} \rceil$ and $\#A_n(\xi) \leq \lceil \frac{1}{\xi} \rceil$.

Put $r(\xi) = [\xi^\xi(1-\xi)^{1-\xi}]^{-1}$. Then we have:

$$(i) \quad P_\theta^n(A_n(\xi) \not\subseteq F_\theta(\eta)) \leq \eta^{-1} [r(\xi)\eta]^n$$

when $\eta \leq \xi$ and $n \geq \xi^{-1}$.

Proof of (i):

$$\begin{aligned} P_\theta^n(A_n(\xi) \not\subseteq F_\theta(\eta)) &\leq \sum_{\theta'} P_\theta^n(\theta' \in A_n(\xi); \theta' \notin F_\theta(\eta)) \\ &= \sum' P_\theta^n(h_n(\theta') \geq \xi) \quad \text{where } \sum' \end{aligned}$$

indicates that the summation is over all points θ' such that $Q_\theta(\theta') < \eta$. Considering a particular term $P_\theta^n(h_n(\theta') \geq \xi)$ we find successively:

$$\begin{aligned} P_\theta^n(h_n(\theta') \geq \xi) &= P_\theta^n(e^{nth_n(\theta')} \geq e^{nt\xi}) \\ &\leq e^{-nt\xi} E_\theta e^{nth_n(\theta')} = [e^{-t\xi} E_\theta e^{th_1(\theta')}]^n; \quad t \geq 0. \end{aligned}$$

Minimizing the last term w.r.t. $t \in [0, \infty[$ we get:

$$\begin{aligned} P_\theta^n(h_n(\theta') \geq \xi) &\leq [(1 - Q_\theta(\theta'))^{1-\xi} Q_\theta(\theta')^\xi r(\xi)]^n \\ &\leq Q_\theta(\theta')^{n\xi} r(\xi)^n = Q_\theta(\theta') Q_\theta(\theta')^{n\xi-1} r(\xi)^n \\ &\leq Q_\theta(\theta') \eta^{n\xi-1} r(\xi)^n \quad \text{when } n\xi \geq 1. \end{aligned}$$

Hence:

$$\begin{aligned} \sum' P_\theta^n(h_n(\theta') \geq \xi) &\leq Q_\theta(F_\theta(\eta)^c) \frac{1}{\eta} [r(\xi)\eta]^n \\ &\leq \frac{1}{\eta} [r(\xi)\eta]^n \quad \text{when } n \geq \xi^{-1}. \end{aligned}$$

Thus (i) is proved.

Consider next the probability that $A_n(\xi)$ does not cover true the value of θ . As above we get successively:

$$\begin{aligned} P_{\theta}^n(\theta \notin A_n(\xi)) &= P_{\theta}^n(h_n(\theta) < \xi) \\ &= P_{\theta}^n(e^{nt(1-h_n(\theta))} > e^{nt(1-\xi)}) \\ &\leq e^{-nt(1-\xi)} E_{\theta} e^{nt(1-h_n(\theta))} = \left[e^{-t(1-\xi)} E_{\theta} e^{t(1-h_1(\theta))} \right]^n, \end{aligned}$$

when $t \geq 0$.

Assuming $\xi \leq Q_{\theta}(\theta)$ and minimizing the last term w.r.t. $t \in [0, \infty[$ we find

$$(ii) \quad P_{\theta}^n(\theta \notin A_n(\xi)) \leq \left[(1-Q_{\theta}(\theta))^{1-\xi} Q_{\theta}(\theta)^{\xi} r(\xi) \right]^n; \quad \xi \leq Q_{\theta}(\theta).$$

Consider so the experiment $\mathcal{E}_{\theta, \eta} = (P_{\theta'}, \theta' \in F_{\theta}(\eta))$ where $\theta \in \Theta$ and $\eta \in]0, 1[$. Let $t_{n, \theta}$ be any maximum likelihood estimator of θ' w.r.t. the experiment $\mathcal{E}_{\theta, \eta}$ and the observations X_1, X_2, \dots, X_n . Suppose $\eta < \alpha$. Then $\theta \in F_{\theta}(\eta)$ so that :

$$\begin{aligned} P_{\theta}^n(t_{n, \theta} \neq \theta) &= \Sigma \{ P_{\theta}^n(t_{n, \theta} = \theta') : \theta' \in F_{\theta}(\eta) - \{\theta\} \} \\ &\leq \Sigma \{ P_{\theta}^n \wedge P_{\theta'}^n : \theta' \in F_{\theta}(\eta) - \{\theta\} \} \\ &\leq \Sigma \{ C(\mathcal{E}_{\{\theta, \theta'\}})^n : \theta' \in F_{\theta}(\eta) - \{\theta\} \}. \end{aligned}$$

The first " \leq " here may be obtained as follows:

$$\begin{aligned} P_{\theta}^n(t_{n, \theta} = \theta') &\leq P_{\theta}^n(f_{\theta', n} \geq f_{\theta, n}) = \int_{[f_{\theta', n} \geq f_{\theta, n}]} f_{\theta, n} = \\ &= \int_{[f_{\theta', n} \geq f_{\theta, n}]} f_{\theta, n} \wedge f_{\theta', n} \leq \int f_{\theta, n} \wedge f_{\theta', n} = \|P_{\theta}^n \wedge P_{\theta'}^n\|. \end{aligned}$$

The last " \leq " follows from corollary 6.23, with $\lambda_1 = \lambda_2 = \frac{1}{2}$ and from corollary 6.22.

Let s_n be any maximum likelihood estimator based on X_1, X_2, \dots, X_n and $A_n(\xi)$. More precisely: s_n is any measurable function of X_1, \dots, X_n such that

$$f_{s_n, n}(x_1, \dots, x_n) \geq f_{\theta', n}(x_1, \dots, x_n)$$

when $\theta' \in A_n(\xi)$ and $s_n = s_n(x_1, \dots, x_n)$. Thus s_n is a restricted M.L. estimator.

Suppose now that $\theta \in A_n(\xi) \subseteq F_\theta(\eta)$. We may then by the slight arbitrariness of $t_{n, \theta}$, require that $t_{n, \theta} = s_n$ whenever $f_{s_n, n}$ is a possible candidate for maximum likelihood in $\mathcal{G}_{\theta, \eta}$. It follows that the implication " $s_n \neq \theta$ " \Rightarrow " $t_{n, \theta} \neq \theta$ " holds when $\theta \in A_n(\xi) \subseteq F_\theta(\eta)$. Hence, for $\eta \leq \xi \leq Q_\theta(\theta)$ and $n \geq \xi^{-1}$:

$$\begin{aligned} P_\theta^n(s_n \neq \theta) &\leq P_\theta^n(\theta \notin A_n(\xi)) + P_\theta^n(A_n(\xi) \not\subseteq F_\theta(\eta)) \\ &+ P_\theta^n(t_{n, \theta} \neq \theta) \leq [(1 - Q_\theta(\theta))^{1-\xi} Q_\theta(\theta)^\xi r(\xi)]^n + \\ &+ \eta^{-1} [r(\xi) \eta]^n + \Sigma \{C(\mathcal{G}_{\{\theta', \theta\}})^n : \theta' \neq \theta \quad Q_\theta(\theta') \geq \eta\}. \end{aligned}$$

Hence; since $Q_\theta(\theta) \geq \alpha$:

$$\delta(\mathcal{G}^n, \mathcal{M}_a)/2 \leq [(1-\alpha)^{1-\xi} \alpha^\xi r(\xi)]^n + \eta^{-1} [r(\xi) \eta]^n + \left[\frac{1}{\eta}\right] C(\mathcal{G})^n$$

when $n \geq \xi^{-1}$, so that

$$\limsup_n \sqrt[n]{\delta(\mathcal{G}^n, \mathcal{M}_a)/2} \leq [(1-\alpha)^{1-\xi} \alpha^\xi r(\xi)] \vee [r(\xi) \eta] \vee C(\mathcal{G})$$

$\eta \rightarrow 0$ and then $\xi \rightarrow 0$ yield finally:

$$\limsup_n \sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)} \leq (1-\alpha) \vee C(\mathcal{E}) = \frac{\delta(\mathcal{E}, \mathcal{M}_a)}{2} \vee C(\mathcal{E}).$$

We proved this inequality under the assumption that $\delta(\mathcal{E}, \mathcal{M}_a) < 2$.

If $\delta(\mathcal{E}, \mathcal{M}_a) = 2$, however, then the inequality holds trivially.

Substituting \mathcal{E}^r , where r is a positive integer, for \mathcal{E} in this inequality we get:

$$\limsup_n \sqrt[n]{\delta(\mathcal{E}^{nr}, \mathcal{M}_a)} \leq C(\mathcal{E})^r \vee \delta(\mathcal{E}^r, \mathcal{M}_a)/2$$

so that

$$\limsup_n \sqrt[nr]{\delta(\mathcal{E}^{nr}, \mathcal{M}_a)} \leq C(\mathcal{E}) \vee \sqrt[r]{\delta(\mathcal{E}^r, \mathcal{M}_a)/2}.$$

For any $n \geq r$ there is a unique $m \geq 1$ so that $mr \leq n \leq (m+1)r$.

Then:

$$\sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)} \leq \sqrt[n]{\delta(\mathcal{E}^{mr}, \mathcal{M}_a)}.$$

The right hand side of this inequality, is between

$$\sqrt[mr]{\delta(\mathcal{E}^{mr}, \mathcal{M}_a)} \quad \text{and} \quad \sqrt[(m+1)r]{\delta(\mathcal{E}^{mr}, \mathcal{M}_a)}.$$

It follows that:

$$\limsup_n \sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)} \leq C(\mathcal{E}) \vee \sqrt[r]{\delta(\mathcal{E}^r, \mathcal{M}_a)/2}.$$

Hence

$$\limsup_n \sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)} \leq C(\mathcal{E}) \vee \tau(\mathcal{E}) = \sigma(\mathcal{E}).$$

On the other hand, for $\theta_1 \neq \theta_2$

$$\liminf_n \sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)} \geq \liminf_n \sqrt[n]{\delta(\mathcal{E}_{\{\theta_1, \theta_2\}}^n)}$$

Hence, by theorem 6.16,

$$\liminf_n \sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)} \geq C(\mathcal{E}_{\{\theta_1, \theta_2\}})$$

so that

$$\liminf_n \sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)} \geq c(\mathcal{E}) .$$

Hence, since the inequality: $\liminf_n \sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)} \geq \tau(\mathcal{E})$ is trivial:

$$\liminf_n \sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)} \geq c(\mathcal{E}) \vee \tau(\mathcal{E}) = \sigma(\mathcal{E}) .$$

Thus

$$\sigma(\mathcal{E}) \leq \liminf_n \sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)} \leq \limsup_n \sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)} \leq \sigma(\mathcal{E})$$

so that

$$(iii) \quad \lim_n \sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)} = \sigma(\mathcal{E}) .$$

The two statements on estimators are, together, merely a rephrasing of (iii). Note, however, that the proof indicates how an asymptotically optimal estimator could be constructed.

□

Corollary 8.2

$$\sigma(\mathcal{E}^r) = \sigma(\mathcal{E})^r ; \quad r = 1, 2, \dots .$$

Proof:

$$\begin{aligned} \sigma(\mathcal{E}^r) &= \lim_n \sqrt[n]{\delta((\mathcal{E}^r)^n, \mathcal{M}_a)} = \lim_n \left[\sqrt[nr]{\delta(\mathcal{E}^{nr}, \mathcal{M}_a)} \right]^r = \\ &= \left[\lim_n \sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)} \right]^r = \sigma(\mathcal{E})^r . \end{aligned}$$

□

Corollary 8.3

Let $\mathcal{E} = (P_\theta : \theta \in \Theta)$ be any experiment with a countable parameterset Θ . Then the following conditions are equivalent:

- (i) $\delta(\mathcal{E}^n, \mathcal{M}_a) < 1$ for some n
- (ii) $\lim_{n \rightarrow \infty} \delta(\mathcal{E}^n, \mathcal{M}_a) = 0$

- (iii) $\delta(\mathcal{E}^n, \mathcal{M}_a) \leq c\rho^n$; $n = 1, 2, \dots$ for some constant $c > 0$ and some constant $\rho < 1$.
- (iv) $\delta(\mathcal{E}^n, \mathcal{M}_a) < 2$ for some n
and $\inf_{\theta_1 \neq \theta_2} \|P_{\theta_1} - P_{\theta_2}\| > 0$.
-

Remark 1

It follows that if $\mathcal{E}^n \rightarrow \mathcal{M}_a$ then the speed of convergence is, provided $\mathcal{E} \nmid \mathcal{M}_a$, necessarily exponential.

Remark 2

The constant 1 in (i) can not be increased. If, for example, $\mathcal{E} = (P_1, P_2, P_3, \dots)$ where $P_1 = P_2$ and $P_i \wedge P_j = 0$ when $i, j \geq 2$ then $\delta(\mathcal{E}^n, \mathcal{M}_a) = 1$ for all $n = 1, 2, \dots$.

Proof:

Suppose $\delta(\mathcal{E}^r, \mathcal{M}_a) < 1$. Then $\tau(\mathcal{E}) \leq \sqrt[r]{\delta(\mathcal{E}^r, \mathcal{M}_a)} < 1$ and $\delta_2(\mathcal{E}^r, \mathcal{M}_a) \leq \delta(\mathcal{E}^r, \mathcal{M}_a) < 1$. It follows

that: $\sup_{\theta_1 \neq \theta_2} \delta(\mathcal{E}_{\{\theta_1, \theta_2\}}^r, \mathcal{M}_a) < 1$, or, equivalently,

that: $\inf_{\theta_1 \neq \theta_2} \delta(\mathcal{M}_i, \mathcal{E}_{\{\theta_1, \theta_2\}}^r) > 0$ i.e.

$C(\mathcal{E})^r = \sup_{\theta_1 \neq \theta_2} C(\mathcal{E}_{\{\theta_1, \theta_2\}}^r) < 1$. Hence, as $\tau(\mathcal{E}) < 1$ and

$C(\mathcal{E}) < 1$; $\sigma(\mathcal{E}) < 1$. Thus, by the theorem, (i) \Rightarrow (iii).

The other implications of the corollary are then straight forward.



Example 8.4 (Continuation of example 3.12)

Consider an experiment \mathcal{E}_ξ as defined in example 3.12.

Then, since $\mathcal{E}_\xi^n = \mathcal{E}_{\xi^n}$:

$$\frac{2a^n b^n}{a^n + b^n} = H(a^n, b^n) \leq \delta(\mathcal{E}_\xi^n, \mathcal{M}_a) \leq 2H(a^n, b^n) \quad \text{where } a = \sup_{\theta} \xi_{\theta}$$

and $b = \sup_{\theta_1 \neq \theta_2} (\xi_{\theta_1} \wedge \xi_{\theta_2})$. It follows that

$$\sigma(\mathcal{E}_\xi) = \lim_n \sqrt[n]{\delta(\mathcal{E}_\xi^n, \mathcal{M}_a)} = b. \quad \text{Furthermore}$$

$$\int dP_{\theta_1}^{1-t} dP_{\theta_2}^t = \xi_{\theta_1}^{1-t} \xi_{\theta_2}^t \quad \text{so that } C(\mathcal{E}_{\{\theta_1, \theta_2\}}) = \xi_{\theta_1} \wedge \xi_{\theta_2}.$$

Using the fact that $\sqrt[n]{E(X)^n}$ is monotonically increasing in n we find that $\sqrt[n]{\delta(\mathcal{E}_\xi^n, \mathcal{M}_a)/2}$ is monotonically decreasing in n . Hence:

$$\sigma(\mathcal{E}_\xi) = C(\mathcal{E}_\xi) = \tau(\mathcal{E}_\xi)$$

for all experiments \mathcal{E}_ξ .

By a slight modification of the proof of theorem 8.1 we get:

Theorem 8.5

Suppose \mathcal{E}^r is ϵ -deficient w.r.t \mathcal{M}_a and let X_1, X_2, \dots be independent observations of \mathcal{E} . Then there is a sequence $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$ of θ such that:

$$\limsup_{n \rightarrow \infty} \sqrt[n]{P_{\theta}^n(\hat{\theta}_n \neq \theta)} \leq \sup\{C(\mathcal{E}_{\{\theta, \theta'\}}); \theta' \neq \theta\} \vee \sqrt[r]{\frac{\epsilon_{\theta}}{2}}; \theta \in \Theta.$$

Example 8.6

Let $\Theta = \{1, 2, \dots\}$ and let, for each θ , P_{θ} be the uniform distribution on $\{1, 2, \dots, \theta\}$. Then $P_{\theta}(\theta) = \frac{1}{\theta}$ so that \mathcal{E} is

$(2 - \frac{2}{\theta} : \theta \in \Theta)$ deficient w.r.t \mathcal{M}_a . It is easily seen that $\inf_{\theta'} \{ \|P_{\theta'} - P_{\theta}\| ; \theta' \neq \theta \} > 0$ for each θ . Thus for some sequence $\hat{\theta}_n$; $n = 1, 2, \dots$ of estimators of θ , such that $\hat{\theta}_n$ is based on \mathcal{E}^n :

$$\limsup_{n \rightarrow \infty} \sqrt{n/P_{\theta}(\hat{\theta}_n \neq \theta)} < 1.$$

In fact; if X_1, X_2, \dots are independent observations of \mathcal{E} then

$$\sqrt{n/P_{\theta}(\bigvee_{i=1}^n X_i \neq \theta)} = 1 - \frac{1}{\theta}; \quad \theta \in \Theta; \quad n = 1, 2, \dots$$

There are not, however, any sequence $\hat{\theta}_n(X_1, X_2, \dots, X_n)$; $n = 1, 2, \dots$ of estimators of θ such that $\sup_{\theta} P_{\theta}^n(\hat{\theta}_n \neq \theta) \rightarrow 0$. This follows, since $C(\mathcal{E}) = 1$, directly from theorem 8.1. Let us see how $\delta(\mathcal{E}^n, \mathcal{M}_a)$; $n = 1, 2, \dots$ behave in this case. It is easily seen that $\bigvee_{i=1}^n X_i$ is sufficient for \mathcal{E}^n . Let $P_{\theta}^{(n)}$ be the distribution of $\bigvee_{i=1}^n X_i$ under θ . Then $P_{\theta}^{(n)}$ assigns masses $\frac{x^n - (x-1)^n}{\theta^n}$ to $x = 1, 2, \dots, \theta$. It follows

$$\text{that } \left\| \bigvee_{\theta=1}^m P_{\theta}^n \right\| = \left\| \bigvee_{\theta=1}^m P_{\theta}^{(n)} \right\| = \sum_{x=1}^m 1 - (1 - \frac{1}{x})^n \leq 1 + \int_1^m [1 - (1 - \frac{1}{x})^n] dx \leq 1 + \int_1^m \frac{n}{x} dx = 1 + n \log m.$$

Hence $\left\| \frac{1}{m} \bigvee_{\theta=1}^m P_{\theta}^n \right\| \rightarrow 0$ as $m \rightarrow \infty$.

It follows that $0 = \inf_{\lambda} \left\| \bigvee_{\theta} \lambda_{\theta} P_{\theta}^n \right\| = 1 - \frac{\delta(\mathcal{E}^n, \mathcal{M}_a)}{2}$ so that $\delta(\mathcal{E}^n, \mathcal{M}_a) = 2$; $n = 1, 2, \dots$

If Θ is finite, then $\delta(\mathcal{E}, \mathcal{M}_a)$ obtains its maximum value, $\delta(\mathcal{M}_i, \mathcal{M}_a)$, if and only if $\mathcal{E} \sim \mathcal{M}_i$, and then

$$\delta(\mathcal{E}^n, \mathcal{M}_a) \equiv_n \delta(\mathcal{M}_i, \mathcal{M}_a).$$

Here is an example showing that this does not extend to the case of an infinite parameter set:

Example 8.7

Suppose Θ is the set of all pairs (i, j) of integers such that $i < j$, and that $P_\theta(i) = P_\theta(j) = \frac{1}{2}$ when $\theta = (i, j)$. Let X_1, X_2, \dots, X_n be n independent observations of \mathcal{E} . It follows directly from the factorization criterion that $(\bigwedge_i X_i, \bigvee_i X_i)$ is sufficient. Put $\hat{\theta}_n(X_1, \dots, X_n) = (\bigwedge_i X_i, \bigvee_i X_i)$ or $= (0, 1)$ as $\bigwedge_i X_i < \bigvee_i X_i$ or $\bigwedge_i X_i = \bigvee_i X_i$. Then, for $n \geq 2$:

$$P_\theta^n(\hat{\theta}_n(X_1, \dots, X_n) \neq \theta) \leq P_\theta^n(X_1 = X_2 = \dots = X_n = \min \theta) \\ + P_\theta^n(X_1 = X_2 = \dots = X_n = \max \theta) = \frac{1}{2^n} + \frac{1}{2^n} = \frac{1}{2^{n-1}}. \text{ It follows}$$

that

$$\delta(\mathcal{E}^n, \mathcal{M}_a) \leq \frac{4}{2^n} \text{ for all } n.$$

Put $Q_{(i,j)}((i,i)) = Q_{(i,j)}((j,j)) = a = 2^{-n}$ and $Q_{(i,j)}((i,j)) = 1 - 2a$ when $i < j$. Then $(\bigwedge_i X_i, \bigvee_i X_i)$ is distributed according to Q_θ when the X 's are distributed according to P_θ^n . Simple calculations yield:

$$\|\bigvee\{P_{(i,j)}^n : 0 \leq i < j \leq N\}\| = \|\bigvee\{Q_{(i,j)} : 0 \leq i < j \leq N\}\| \\ = (N+1)a + (N+1)N(1-2a) \text{ so that :}$$

$$1 - \|\bigvee\{\frac{1}{N(N+1)} P_{(i,j)}^n : 0 \leq i < j \leq N\}\| \rightarrow 2a$$

as $N \rightarrow \infty$. Hence $\delta(\mathcal{E}^n, \mathcal{M}_a) \geq 4a = \frac{4}{2^n}$. Altogether we have shown that

$$\delta(\mathcal{E}^n, \mathcal{M}_a) \equiv \frac{4}{2^n}.$$

Hence $\delta(\mathcal{E}, \mathcal{M}_a) = 2$ while $\delta(\mathcal{E}^n, \mathcal{M}_a) \leq 1$ when $n \geq 2$.

It is not difficult to see that $C(\mathcal{E}_{\{\theta_1, \theta_2\}}) = \frac{1}{2}$ or $= 0$ as $\theta_1 \cap \theta_2 \neq \emptyset$ or $\theta_1 \cap \theta_2 = \emptyset$ - provided $\theta_1 \neq \theta_2$.

Hence $C(\mathcal{E}) = \tau(\mathcal{E}) = \sigma(\mathcal{E}) = \frac{1}{2}$.

In all the examples we have considered so far the constant $C(\mathcal{E})$ alone determined the rate of exponential convergence. If Θ is finite, then, by section 7, this is always the case. In the infinite case, however, there are other possibilities.

Proposition 8.8

If $\mathcal{E} = (P_\theta : \theta \in \Theta)$ has an accumulation point for the topology of set wise convergence then

$$\delta(\mathcal{E}^n, \mathcal{M}_a) \stackrel{\equiv}{n} 2.$$

This does not, however, exclude the possibility that $C(\mathcal{E}) < 1$.

Remark:

Let μ be a σ -finite measure dominating \mathcal{E} and put $f_\theta = dP_\theta/d\mu$. Then the conditions of the theorem is satisfied whenever Θ is infinite and the densities $f_\theta : \theta \in \Theta$ are uniformly bounded. Thus, in particular $\delta(\mathcal{E}^n, \mathcal{M}_a) \stackrel{\equiv}{n} 2$ when the sample space of \mathcal{E} is finite and Θ is infinite.

By section 7 we have always weak convergence (i.e. convergence for restrictions to finite sub parameter sets) to \mathcal{M}_a provided $P_{\theta_1} \neq P_{\theta_2}$ when $\theta_1 \neq \theta_2$. Thus, if $P_{\theta_1} \neq P_{\theta_2}$ when $\theta_1 \neq \theta_2$ and the conditions of the proposition is satisfied,

then \mathcal{E}^n converges weakly to \mathcal{M}_a although $\delta(\mathcal{E}^n, \mathcal{M}_a) \equiv \frac{2}{n}$.

Proof of proposition 8.8:

Let P be an accumulation point of $(P_\theta: \theta \in \Theta)$ for the topology of set wise convergence of measures. Then, since Θ is countable, there is a sequence $\theta_1, \theta_2, \dots$ of distinct elements of Θ such that $P_{\theta_n} \neq P$ for all n and such that P_{θ_n} converges, as $n \rightarrow \infty$, to P for this topology. Then, by the Vitali-Hahn-Saks theorem, P is a probability measure. Put $\Theta_0 = \{\theta_1, \theta_2, \dots\}$. By the Vitali-Hahn-Saks theorem again, $P_{\theta_n}^r$ converges to P^r for the topology of set wise convergence for \mathcal{E}^r . Let M be any randomized estimator of $\theta \in \Theta_0$ in $\mathcal{E}_{\Theta_0}^r$ and let $\alpha \leq P_\theta^r(M=\theta)$; $\theta \in \Theta_0$. Let F be a finite subset of Θ_0 . Then, provided n is sufficiently large, $\theta_n \notin F$. Moreover, if $\theta_n \notin F$ then:

$$P_{\theta_n}^r(M \in F) \leq P_{\theta_n}^r(M \neq \theta_n) \leq 1 - \alpha.$$

Hence; by letting $n \rightarrow \infty$:

$$P^r(M \in F) \leq 1 - \alpha \quad \text{for all}$$

finite subsets F of Θ_0 . It follows that $1 = P^r(M \in \Theta_0) = \sup_F P^r(M \in F) \leq 1 - \alpha$, so that $\alpha \leq 0$. Hence:

$$\inf_{\Theta_0} P_\theta^r(M=\theta) = 0$$

for all randomized estimators of θ in $\mathcal{E}_{\Theta_0}^r$.

Thus $2 \geq \delta(\mathcal{E}^r, \mathcal{M}_a) \geq \delta(\mathcal{E}_{\Theta_0}^r, \mathcal{M}_a) \geq 2$. This proves the first statement and the last statement follows from the example below.

Example 8.9

Suppose $\Theta = \{1, 2, \dots\}$ and that P_θ , $\theta = 1, 2, \dots$, is the probability measure on $[0, 1]$ whose density f_θ w.r.t. the

uniform distribution P is:

$$f_{\theta}(x) = 2^{-\theta} \sum_{k=1}^{\infty} I\left[\frac{k-1}{2^{\theta}}, \frac{k-1}{2^{\theta}}\right] \cdot \quad \text{Then}$$

$$\begin{aligned} f_{\theta}(x) &= 2^{-\theta} \sum_{k=1}^{\infty} I\left[\frac{(k-1)2^{\eta}}{2^{\theta+\eta}}, \frac{(k-1)2^{\eta}}{2^{\eta+\theta}}\right] \\ &= 2^{-\theta} \sum_{k=1}^{\infty} \sum_{j=(k-1)2^{\eta}+1}^{(k-1)2^{\eta}} I\left[\frac{j-1}{2^{\theta+\eta}}, \frac{j}{2^{\theta+\eta}}\right] \end{aligned}$$

It follows that:

$$\int f_{\theta}^{1-t} f_{\theta+\eta}^t dP = \sum_{k=1}^{\infty} \sum_{j=(k-1)2^{\eta}+1}^{(k-1)2^{\eta}} 2^{1-t} 2^t \frac{1}{2} \cdot \frac{1}{2^{\theta+\eta}} = \frac{1}{2}$$

when $\eta \geq 1$.

Thus any dichotomy (P_{θ}, P_{η}) where $\theta \neq \eta$ is equivalent to the simple dichotomy

$x \backslash$	0	1	2
θ	$\frac{1}{2}$	0	$\frac{1}{2}$
η	0	$\frac{1}{2}$	$\frac{1}{2}$

It follows that $C(\mathcal{G}) = \frac{1}{2}$ and that $\|P_{\theta} - P_{\eta}\| = 1$ when $\theta \neq \eta$.

Consider next any number $t \in]0, 1[$. Let, for each $n = 1, 2, \dots$, i be chosen so that:

$$\frac{i-1}{2^n} \leq t < \frac{i}{2^n}.$$

Then:

$$t - \frac{1}{2^n} < \frac{i-1}{2^n} = P_n[0, \frac{i-1}{2^n}[\leq P_n[0, t[\leq P_n[0, \frac{i}{2^n}[= \frac{i}{2^n} \leq t + \frac{1}{2^n}.$$

It follows that:

$$P_n[0, t[\rightarrow t; \quad t \in]0, 1[$$

so that

$$P_n \rightarrow P$$

in the weak topology of $C[0,1]^*$.

Let, more generally, B be any Borel subset of $[0,1]$, and let τ be any limit point of the sequence $P_1(B), P_2(B), \dots$. Then there are integers $\theta_1 < \theta_2 < \dots$ so that $P_{\theta_i}(B) \rightarrow \tau$ as $i \rightarrow \infty$. We may, by the uniform integrability of f_θ ; $\theta = 1, 2, \dots$, assume that there is a P integrable function g on $[0,1]$ such that $\int h f_{\theta_i} dP \rightarrow \int h g dP$ for any $h \in L_1(P)$. It follows in particular that $\tau = \int_B g dP$ and that $t = \int_0^t g dP$; $t \in]0, 1[$. Hence $g = 1$ a.e. P so that $\tau = P(B)$. Thus P_θ converges to P , as $\theta \rightarrow \infty$, for the topology of set wise convergence. Hence by proposition 8.7;

$$\delta(\mathcal{E}^n, \mathcal{M}_a) = 2; \quad n = 1, 2, \dots$$

although $C(\mathcal{E}) = \frac{1}{2}$. This example shows that the constant $C(\mathcal{E})$ alone does not, in general, determine the rate of exponential convergence.

Here is an example, strengthening the last example, showing that we may have $C(\mathcal{E}) < \tau(\mathcal{E})$ for arbitrarily small values of $\tau(\mathcal{E})$.

Example 8.10

Let Θ and $(P_\theta : \theta \in \Theta)$ be defined as in the previous example. Choose a constant $a \in [0, 1]$ and put, for each θ , $\tilde{P}_\theta = (1-a)P_\theta + aQ_\theta$ where Q_θ is the one point distribution in θ . Let X_1, \dots, X_n be n independent observations of $\tilde{\mathcal{E}} = (\tilde{P}_\theta : \theta \in \Theta)$. Put $\hat{\theta}_n(X_1, \dots, X_n) = \bigvee_i X_i \vee 1$.

Then

$$P_{\theta}^n(\hat{\theta}_n \neq \theta) \leq (1-a)^n; \quad n = 1, 2, \dots \quad \text{so}$$

that

$$\delta(\mathcal{E}^n, \mathcal{M}_a) \leq 2(1-a)^n; \quad n = 1, 2, \dots$$

Let $\check{\theta}(X_1, \dots, X_n)$ be any estimator of θ and consider independent variables $(Y_1, Z_1), (Y_2, Z_2), \dots, (Y_n, Z_n)$ such that, for each θ ,

$$(i) \quad \Pr(Y_k = 0) = 1-a \quad \text{and} \quad \Pr(Y_k = 1) = a; \quad k = 1, \dots, n.$$

$$(ii) \quad \text{The conditional distribution of } Z_k \text{ given } Y_k = 0 \text{ is } P_{\theta}.$$

$$(iii) \quad \text{The conditional distribution of } Z_k \text{ given } Y_k = 1 \text{ is } Q_{\theta}.$$

Then $\mathcal{L}(Z_1, \dots, Z_n | \theta) = \mathcal{L}(X_1, \dots, X_n | \theta)$. Hence

$$\begin{aligned} \tilde{P}_{\theta}^n(\check{\theta}(X_1, \dots, X_n) \neq \theta) &= \Pr(\check{\theta}_n(Z_1, \dots, Z_n) \neq \theta | \theta) \\ &= \sum_y (1-a)^{n-\Sigma Y} a^{\Sigma Y} P_r(\check{\theta}_n(Z_1, \dots, Z_n) \neq \theta | \theta, Y=y) \\ &\geq (1-a)^n P(\check{\theta}_n(Z_1, \dots, Z_n) \neq \theta | \theta, Y=0) \\ &= (1-a)^n P_{\theta}^n(\check{\theta}_n(X_1, \dots, X_n) \neq \theta), \text{ so that:} \end{aligned}$$

$$\sup_{\theta} \tilde{P}_{\theta}^n(\check{\theta}_n(X_1, \dots, X_n) \neq \theta) \geq (1-a)^n \sup_{\theta} P_{\theta}^n(\check{\theta}_n(X_1, \dots, X_n) \neq \theta)$$

and the last quantity is, by the previous example, at least equal to $(1-a)^n$. Hence $\delta(\mathcal{E}^n, \mathcal{M}_a) \geq 2(1-a)^n$. Altogether we have shown that:

$$\delta(\mathcal{E}^n, \mathcal{M}_a) \equiv_n 2(1-a)^n.$$

It follows, in particular, that $\tau(\mathcal{E}) = 1-a$.

On the other hand, by example 8.8 again:

$$\int d\tilde{P}_\theta^{1-t} d\tilde{P}_\eta^t = \frac{1}{2} \int dP_\theta^{1-t} dP_\eta^t = \frac{1}{2}(1-a) \quad \text{when } \theta \neq \eta.$$

Thus $C(\mathcal{E}) = \frac{1}{2}(1-a)$.

One might conjecture, on the basis of our examples, that $\sqrt[n]{\delta(\mathcal{E}^n, \mathcal{M}_a)/2}$ is monotonically decreasing in n when Θ is infinite. That this, however, is not the case, may be seen from theorem 6.16, and the fact that any sequence $\delta(\mathcal{E}^n, \mathcal{M}_a)$; $n = 1, 2, \dots$ where \mathcal{E} is a dichotomy may be realized as a sequence $\delta(\mathcal{F}^n, \mathcal{M}_a)$; $n = 1, 2, \dots$ for an experiment \mathcal{F} with countably infinite parameterset and having the same C value as \mathcal{E} has. If $\mathcal{E} = (P_1, P_2)$ then $\delta(\mathcal{F}^n, \mathcal{M}_a) \equiv \delta(\mathcal{E}^n, \mathcal{M}_a)$ provided $\mathcal{F} = (P_1, P_2, Q_3, Q_4, \dots)$ where $(P_1 + P_2) \wedge Q_i = Q_i \wedge Q_j = 0$ when $i \neq j$ and $i, j \geq 3$.

Consider again a sequence X_1, X_2, \dots of independent observations of the experiment $\mathcal{E} = (P_\theta; \theta \in \Theta)$. If $\hat{\theta}_n(X_1, X_2, \dots, X_n)$; $n = 1, 2, \dots$ is a sequence of estimators of θ then we put:

$$N_\theta(\hat{\theta}_1, \hat{\theta}_2, \dots) = \max\{n: \hat{\theta}_n \neq \theta\}$$

if $\hat{\theta}_n = \theta$ for n sufficiently large. If, on the other hand $\hat{\theta}_n \neq \theta$ for arbitrarily large n , then we put $N_\theta(\hat{\theta}_1, \hat{\theta}_2, \dots) = \infty$. Then we have

Theorem 8.11

$\mathcal{E}^n \rightarrow \mathcal{M}_a$ if and only if there is a sequence $\hat{\theta}_1, \hat{\theta}_2, \dots$ of estimators of θ such that $(\mathcal{L}_\theta(N_\theta(\hat{\theta}_1, \hat{\theta}_2, \dots)); \theta \in \Theta)$ is tight. If so, then $\hat{\theta}_1, \hat{\theta}_2, \dots$ may be chosen so that

$$x \rightsquigarrow e^{tx}$$

is uniformly integrable w.r.t. $\mathcal{L}_\theta(N_\theta(\hat{\theta}_1, \hat{\theta}_2, \dots)) : \theta \in \Theta$ when $t > 0$ is sufficiently small.

Proof: Let $\hat{\theta}_n(X_1, X_2, \dots, X_n)$ be such that $\sup_\theta P_\theta^n(\hat{\theta}_n \neq \theta) = \frac{1}{2} \delta(\mathcal{C}^n, \mathcal{M}_a)$. Suppose first that $\mathcal{C}^n \rightarrow \mathcal{M}_a$. Then, by corollary 8.3: $\sqrt[n]{\delta(\mathcal{C}^n, \mathcal{M}_a)} \rightarrow \rho_1$ where $\rho_1 < 1$. Thus

$\sqrt[n]{\sup_\theta P_\theta^n(\hat{\theta}_n \neq \theta)} \rightarrow \rho_1$. Let $\rho_1 < \rho_2 < 1$. Then

$\sup_\theta P_\theta^n(\hat{\theta}_n \neq \theta) \leq \rho_2^n$ when n is sufficiently large, $n \geq n_0$ say.

It follows that

$$P_\theta^\infty(N_\theta \geq k) \leq \sum_{n=k}^{\infty} P_\theta^\infty(\hat{\theta}_n \neq \theta) \leq \sum_{n=k}^{\infty} \rho_2^n \leq \epsilon$$

for all θ , provided k is sufficiently large. Hence, by partial integration, $x \rightsquigarrow e^{tx}$, is uniformly integrable w.r.t. $\mathcal{L}_\theta(N_\theta)$; $\theta \in \Theta$ when $t < \log \rho_1^{-1}$. This proves, of course, also that $(\mathcal{L}_\theta(N_\theta) : \theta \in \Theta)$ is tight. Suppose conversely that $(\mathcal{L}_\theta(N_\theta) : \theta \in \Theta)$ is tight for some sequence $\hat{\theta}_n$; $n = 1, 2, \dots$. Let $\epsilon > 0$. Then there is a k_ϵ so that $P_\theta(N_\theta \geq k) \leq \epsilon$ for $k \geq k_\epsilon$ for all θ . Hence $P_\theta(\hat{\theta}_n \neq \theta) \leq \epsilon$ when $n \geq k_\epsilon$, so that $\delta(\mathcal{C}^n, \mathcal{M}_a) \leq 2\epsilon$ for $k \geq k_\epsilon$. The theorem follows now from corollary 8.3. □

The assumption that $\mathcal{L}_{P_\theta}(N_\theta)$; $\theta \in \Theta$ is tight is essential. There is always a sequence $\hat{\theta}_n(X_1, X_2, \dots, X_n)$; $n = 1, 2, \dots$ such that $P_\theta(N_\theta(\hat{\theta}_1, \hat{\theta}_2, \dots) < \infty) \equiv 1$ provided each parameter point θ is isolated for statistical distance and \mathcal{C} is ϵ -deficient w.r.t. \mathcal{M}_a for some function ϵ such that $\epsilon_\theta < 2$ for all θ . If this hold, then $(\hat{\theta}_1, \hat{\theta}_2, \dots)$ may be chosen so that, for each θ , $E_\theta e^{tN_\theta} < \infty$ when $t > 0$ is sufficiently

small. The required smallness may, however, depend on θ .

As in section 6 and 7 the exponential convergence of $\delta(\mathcal{E}^n, \mathcal{M}_a)$ to zero imply the exponential convergence to zero of many interesting functionals - in particular functionals associated with decision problems. In many cases, however, the upper bounds obtained from $\delta(\mathcal{E}, \mathcal{M}_a)$ alone is too crude to be of interest. Consider, again, the problem of catching θ with an r -point confidence set. The minimax probability of not catching θ is then:

$$\kappa_r = 1 - \inf_{\lambda} \left\| \bigvee_U \sum_{\theta \in U} \lambda_{\theta} P_{\theta} \right\|$$

where U runs through all r -point subsets of Θ . Then $\kappa_1 = \delta(\mathcal{E}, \mathcal{M}_a)/2$ and it is easily seen that κ_r is monotonically decreasing in r . If P_{θ} is, for each integer θ , the uniform-distribution on $(\theta, \theta+1)$ then, by the next section, $\kappa_1 = 1$ although it is trivial that $\kappa_3 = 0$.

In example 8.6 we have^{*)} $t_r(\mathcal{E}^n)_{r,n} \equiv 1$ although:

$$P_{\theta}^n(\bigvee_i X_i \leq \theta < (1+\epsilon) \bigvee_i X_i) \geq 1 - \frac{1}{(1+\epsilon)^n}$$

for all θ when $\epsilon > 0$.

We have in this section and the next section chosen to limit ourselves to an investigation of the deviation from total information. Let us, however, note a few simple facts on the deviation from total ignorance.

If $\mathcal{E}^n \rightarrow \mathcal{M}_a$ then, by corollary 8.3 and by the inequalities:

$$2 - \delta(\mathcal{M}_i, \mathcal{E}^n) \leq \delta(\mathcal{E}^n, \mathcal{M}_a), \quad \delta(\mathcal{M}_i, \mathcal{E}^n) \rightarrow 2$$

*) This follows from the inequality: $1 - \kappa_r \leq r(1 - \kappa_1)$,

with exponential speed. In general, by this inequality:

$$\limsup_n \sqrt[n]{2 - \delta(\mathcal{M}_i, \mathcal{E}^n)} \leq \sigma(\mathcal{E}) .$$

It may, however, easily happen that $\delta(\mathcal{M}_i, \mathcal{E}^n) \rightarrow 2$ while $\delta(\mathcal{E}^n, \mathcal{M}_a) \geq 1$ for all n . This is, for example, the case if $\mathcal{E} = (P_1, P_2, \dots)$ where $P_1 = P_2$ and $P_i \wedge P_j = 0$ when $i \neq j$ and $i, j \geq 2$. [Then $\delta(\mathcal{M}_i, \mathcal{E}^n) = 2$; $n = 1, 2, \dots$ while $\delta(\mathcal{E}^n, \mathcal{M}_a) = 1$; $n = 1, 2, \dots$].

Let $\gamma(\theta_1, \theta_2) = \int \sqrt{dP_{\theta_1} dP_{\theta_2}}$ denote the affinity between P_{θ_1} and P_{θ_2} . Then, by corollary 4.2

$$\begin{aligned} 1 - \frac{1}{2} \delta(\mathcal{M}_i, \mathcal{E}) &\geq 1 - \max_{\theta_1 \neq \theta_2} \|P_{\theta_1} - P_{\theta_2}\|/2 = \min_{\theta_1 \neq \theta_2} \|P_{\theta_1} \wedge P_{\theta_2}\| \\ &\geq \frac{1}{4} \max_{\theta_1 \neq \theta_2} \gamma(\theta_1, \theta_2)^2 . \end{aligned}$$

Applying this to \mathcal{E}^n , and using the multiplicativity of the affinity, we find that:

$$\liminf_n \sqrt[n]{1 - \frac{1}{2} \delta(\mathcal{M}_i, \mathcal{E}^n)} \geq \inf_{\theta_1 \neq \theta_2} \gamma(\theta_1, \theta_2)^2 .$$

Altogether we have proved:

Proposition 8.12

$$\inf_{\theta_1 \neq \theta_2} \gamma(\theta_1, \theta_2)^2 \leq \liminf_{n \rightarrow \infty} \sqrt[n]{1 - \frac{1}{2} \delta(\mathcal{M}_i, \mathcal{E}^n)} \leq \limsup_{n \rightarrow \infty} \sqrt[n]{1 - \frac{1}{2} \delta(\mathcal{M}_i, \mathcal{E}^n)} \leq \sigma(\mathcal{E}) .$$

9. Replicated translation experiments on the integers.

Let, for each distribution P on the integers and each integer θ , P_θ denote the right θ -translate of the distribution P .
Thus :

$$\mathcal{L}(X + \theta) = P_\theta$$

when

$$\mathcal{L}(X) = P.$$

The experiment $(P_\theta; \theta \in \Theta)$ will be denoted by \mathcal{E}_P .
It was shown in [20] that :

$$\delta(\mathcal{E}_P^n, \mathcal{M}_a) = 2(1 - \sum_{x_2, \dots, x_n} \Gamma(x_2, \dots, x_n))$$

where, for each $(n-1)$ tuple (x_2, \dots, x_n) of integers:

$$\Gamma(x_2, \dots, x_n) = \max_x P^n(x + x_2, \dots, x + x_n).$$

Minimax estimators may be found by :

Theorem 9.1.

Let X_1, \dots, X_n be n independent observations of \mathcal{E} . Then there are translation invariant maximum likelihood estimators of θ and any such estimator $\hat{\theta}_n$ is minimax for the problem of guessing θ with 0-1 loss, i.e. :

$$P_\theta(\hat{\theta}_n \neq \theta) = \delta(\mathcal{E}_P^n, \mathcal{M}_a)/2.$$

Proof :

Let $b(x_2, \dots, x_n)$ maximize $P^n(-\theta, x_2 - \theta, \dots, x_n - \theta)$ w.r.t. θ and put $\hat{\theta}_n(X_1, X_2, \dots, X_n) = X_1 + b(X_2 - X_1, \dots, X_n - X_1)$. Then $\hat{\theta}_n$ is translation invariant and :

$$\begin{aligned} P_{\hat{\theta}_n}^n(X_1, \dots, X_n) &= P^n(X_1 - \hat{\theta}_n, \dots, X_n - \hat{\theta}_n) = \\ &= P^n(-\hat{\theta}_n(0, X_2 - X_1, \dots, X_n - X_1), X_2 - X_1 - \hat{\theta}_n(0, X_2 - X_1, \dots), \dots) \\ &= P^n(-b(X_2 - X_1, \dots, X_n - X_1), X_2 - X_1 - b(X_2 - X_1, \dots), \dots) \\ &\geq P^n(-\theta, X_2 - X_1 - \theta, X_3 - X_1 - \theta, \dots) ; \theta \in \Theta . \end{aligned}$$

Hence, by substituting $X_1 - \theta$ for $-\theta$:

$P_{\hat{\theta}_n}^n(X_1, \dots, X_n) \geq P_{\theta}^n(X_1, \dots, X_n)$ so that $\hat{\theta}_n$ is a maximum likelihood estimator of θ . Consider next any translation invariant maximum likelihood estimator $\tilde{\theta}_n(X_1, \dots, X_n)$. Then $\tilde{\theta}_n(X_1, \dots, X_n)$ assigns mass 1 to the set $\{\theta : P_{\theta}^n(X_1, \dots, X_n) = \max_{\theta'} P_{\theta'}^n(X_1, \dots, X_n)\}$.

By translation invariance :

$$\begin{aligned} P_{\theta}^n(\tilde{\theta}_n \neq \theta) &= P^n(\tilde{\theta}_n \neq 0) = 1 - P^n(\tilde{\theta}_n = 0) \text{ and} \\ P^n(\tilde{\theta}_n = 0) &= \sum P^n(\tilde{\theta}_n = 0 | X=x) P^n(x) \\ &= \sum' P^n(\tilde{\theta}_n = -x_1 | X_1=0, X_2=x_2-x_1, \dots) P^n(x) \end{aligned}$$

where \sum' indicates that the summation is over all n -tuples (x_1, \dots, x_n) so that $P^n(x) = \max_{\theta} P_{\theta}^n(x)$.

Substituting $y_i = x_i - x_1$; $i \geq 2$ we get :

$$P^n(\tilde{\theta}_n = 0) = \sum_y \sum_{x_1}'' P^n(\tilde{\theta}_n = -x_1 | X_1=0, X_2=y_2, \dots, X_n=y_n) \max_{\theta} P^n(-\theta, y_2-\theta, \dots)$$

where Σ'' indicates that the summations is over all x_1 such that $P_{-x_1}^n(0, y_2, \dots, y_n) \geq P_{\theta}^n(0, y_2, \dots, y_n)$ for all θ . Hence, since $\tilde{\theta}_n$ is a maximum likelihood estimator :

$$P^n(\tilde{\theta}_n = 0) = \sum_y \max_{\theta} P^n(-\theta, y_2-\theta, \dots) = \sum_y \Gamma(y)$$

so that : $P_{\theta}^n(\tilde{\theta}_n \neq \theta) = 1 - \sum_y \Gamma(y) = \delta(\mathcal{C}^n, \mathcal{U}_a)/2$.



Corollary 9.2.

Suppose, for some integer a , $P(x) = 0$ when $x < a$. Then $\delta(\mathcal{C}^n, \mathcal{U}_a) \leq 2(1-P(a))^n$ and " $=$ " holds provided $P(a) \geq P(a+1) > P(a+2) \geq \dots$

Remark.

Suppose $P(a) \geq P(a+1) \geq \dots$. Let us compute $C(\mathcal{C})$ in this case. Suppose $a=0$. Then :

$$\begin{aligned} C(\mathcal{C}_{\{\theta_1, \theta_2\}}) &= C(\mathcal{C}_{\{0, |\theta_2 - \theta_1|\}}) = \inf_{0 < t < 1} \sum_{x \geq |\theta_2 - \theta_1|} P(x) \left[\frac{P(x - |\theta_2 - \theta_1|)}{P(x)} \right]^t \\ &= \lim_{t \rightarrow 0} \sum_{x \geq |\theta_2 - \theta_1|} P(x) \left[\frac{P(x - |\theta_2 - \theta_1|)}{P(x)} \right]^t = \sum_{x \geq |\theta_2 - \theta_1|} P(x) . \end{aligned}$$

Hence $C(\mathcal{C}) = C(\mathcal{C}_{\{0, 1\}}) = \sum_{x \geq 1} P(x) = 1 - P(0)$.

It follows, for arbitrary a , that $C(\mathcal{E}) = 1 - P(a)$. Hence, by the corollary, $C(\mathcal{E}) = \tau(\mathcal{E}) = \sigma(\mathcal{E}) = 1 - P(a)$.

Proof of the corollary :

Put $\hat{\theta}_n(x_1, \dots, x_n) = \bigwedge_{i=1}^n X_i - a$. Then $\hat{\theta}_n$ is translation invariant. Hence :

$P_{\theta}^n(\hat{\theta}_n \neq \theta) = P_0^n(\hat{\theta}_n \neq 0) = (1 - P(a))^n$. This proves " \leq " and the last statement follows by the theorem and by noting that $\hat{\theta}_n$ is a maximum likelihood estimator of θ in this case. □

It follows from the formula $\delta(\mathcal{E}, \mathcal{M}_a)/2 = 1 - \bigvee_x P(x)$ that $\delta(\mathcal{E}, \mathcal{M}_a) < 2$. Furthermore $\inf_{\theta_1 \neq \theta_2} \|P_{\theta_1} - P_{\theta_2}\| = \inf_{\theta \neq 0} \|P_{\theta} - P\| > 0$.

Hence, by theorems 8.1 and 9.1 :

Theorem 9.3.

Let P be a probability distribution on the integers and let, for $n=1, 2, \dots$, $\hat{\theta}_n$ be a translation invariant maximum likelihood estimator of θ based on \mathcal{E}_P^n . Then there are constants $c > 0$ and $\rho \in [0, 1[$ so that :

$$P_{\theta}^n(\hat{\theta}_n \neq \theta) = \frac{\delta(\mathcal{E}_P^n, \mathcal{M}_a)}{2} \leq c \rho^n ; n=1, 2, \dots$$

Theorem 9.3 tells us that the experiment obtained by taking n replications of any experiment \mathcal{E}_P , where P is non degenerate, converges, as $n \rightarrow \infty$, to the totally informative experiment with exponential speed. In spite of this there are experiments \mathcal{E}_P ,

with P non degenerated,

such that two replications are not, in terms of minimax risk, better than one.

Example 9.4.

Put $P(x) = p/a^{|x|}$; $x = \dots, -1, 0, 1, \dots$ where $p \in [0, 1[$ and $a = \frac{3-p}{1-p}$. Let X and Y be independent observations of \mathcal{G}_P . Then, for all integers x and y :

$$P(x) = P(-x) \quad \text{and} \quad P(x+y)P(0) \geq P(x)P(y) .$$

Hence :

$$P_{\theta}^2(x, y) = P(x-\theta)P(y-\theta) = P(\theta-x)P(y-\theta) \leq P(0)P(y-x) = P_x^2(x, y) .$$

It follows that $\hat{\theta}(X, Y) = X$ is a translation invariant maximum likelihood estimator based on \mathcal{G}_P as well as on \mathcal{G}_P^2 . Thus :

$$\delta(\mathcal{G}_P^2, \mathcal{M}_a)/2 = \delta(\mathcal{G}_P, \mathcal{M}_a)/2 = 2(1-p) .$$

It is a fact, almost too trivial to be mentioned, that two or more observations must be at least as good as having one observation, - provided we do not take the cost of observing into consideration. One would also expect that if we are able to do extremely well with two observations, then we should be able to do at least moderately well with one observation. If Θ is finite then this follows immediately from the compactness of Δ -convergence. It may also be substantiated by inequalities like :

Proposition 9.5.

Suppose Θ is a m -point set. Then

$$\delta(\mathcal{G}, \mathcal{M}_a) \leq 2m(m-1) \sqrt{\frac{2n^2}{2\delta(\mathcal{G}^n, \mathcal{M}_a)}} \quad ; \quad n=1, 2, \dots .$$

This inequality may be deduced from proposition 3.8, corollary 6.24 and proposition 6.18 as follows :

$$\begin{aligned}
 \delta(\mathcal{E}, \mathcal{M}_a) &\leq \sum_{\theta_1 \neq \theta_2} \delta(\mathcal{E}_{\{\theta_1, \theta_2\}}, \mathcal{M}_a) \leq 2 \sum_{\theta_1 \neq \theta_2} C(\mathcal{E}_{\{\theta_1, \theta_2\}}) \\
 &\leq 2 \sum_{\theta_1 \neq \theta_2} \gamma(\mathcal{E}_{\{\theta_1, \theta_2\}}) = 2 \sum_{\theta_1 \neq \theta_2} \sqrt{2n [\gamma(\mathcal{E}_{\{\theta_1, \theta_2\}}^n)]^2} \\
 &\leq \sum_{\theta_1 \neq \theta_2} \sqrt{2n^2 \delta(\mathcal{E}_{\{\theta_1, \theta_2\}}^n, \mathcal{M}_a)} \leq 2m(m-1) \sqrt{2n^2 \delta(\mathcal{E}^n, \mathcal{M}_a)} .
 \end{aligned}$$

□

Although this inequality is likely to be very inaccurate in most situations it has one feature which can't, without limitations on \mathcal{E} , be improved. This feature is the role of m , i.e. the cardinality of Θ . We shall here satisfy ourselves with an example of a translation experiment where we may, on the basis of two observations, guess θ with marvelous accuracy while any estimator based on one observation is awfully inaccurate. To make things more concrete we may choose the constant so that the maximum probability of a wrong guess for some estimators based on two observations is less than 10^{-200} while, on the other hand, the supremum of probabilities of making a wrong guess is greater than $1-10^{-200}$ for any estimator based on one observation.

Example 9.6.

Let P be the uniform distribution on some finite set F of integers. Thus, if F contains N integers, then $P(x) = N^{-1}$ or $= 0$ as $x \in F$ or $x \notin F$. Then $\delta(\mathcal{E}_P, \mathcal{M}_a)/2 = 1 - \bigvee_x P(x) = 1 - \frac{1}{N}$. It follows that for any estimator $\hat{\theta}$ based on one observation :

$$\max_{\theta} P_{\theta}(\hat{\theta} \neq \theta) \geq 1 - \frac{1}{N}$$

Suppose n independent observations, X_1, X_2, \dots, X_n of \mathcal{G}_P are available. Clearly $\max_{\theta} P_{\theta}^n(X_1 - \theta, X_2 - \theta, \dots, X_n - \theta) = N^{-n}$ or $= 0$ as $(X_1 - F) \cap (X_2 - F) \cap \dots \cap (X_n - F) \neq \emptyset$ or $= \emptyset$. It follows that a translation invariant maximum likelihood estimator $\hat{\theta}_n$ may, for example, be given by :

$$\hat{\theta}_n = \begin{cases} = \min [X_1 - F) \cap (X_2 - F) \cap \dots \cap (X_n - F)] & \text{if } (X_1 - F) \cap \dots \cap (X_n - F) \neq \emptyset \\ = X_1 & \text{otherwise} \end{cases}$$

We find, successively, that :

$$\begin{aligned} P^n(\hat{\theta}_n = 0) &= \sum \{P^n(x) : \min [(x_1 - F) \cap \dots \cap (x_n - F)] = 0\} \\ &+ \sum \{P^n(x) : (x_1 - F) \cap \dots \cap (x_n - F) = \emptyset\} \\ &= N^{-n} \# \{x : \min [(x_1 - F) \cap \dots \cap (x_n - F)] = 0\} \end{aligned}$$

If $n \geq 2$ then this may be written :

$$\begin{aligned} P^n(\hat{\theta}_n = 0) &= N^{-n} \# \{(x_1, y_2, \dots, y_n) : \min [(x_1 - F) \cap (y_2 + x_1 - F) \cap \dots \\ &\quad \cap (y_n + x_1 - F)] = 0\} \\ &= N^{-n} \# \{(y_2, \dots, y_n) : (y_2, \dots, y_n) \in F^{n-1}\text{-diagonal } (F^{n-1})\} . \end{aligned}$$

It follows that :

$$\delta(\mathcal{G}_P^n, \mathcal{M}_a)/2 \equiv P_{\theta}^n(\hat{\theta}_n \neq \theta) = 1 - N^{-n} \# [F^{n-1}\text{-diagonal } (F^{n-1})]; n=2, 3, \dots$$

The number of elements in $[F^{n-1}\text{-diagonal}(F^{n-1})]$ varies, for fixed n and N , with the structure of the set F . This number is never greater than N^{n-1} and this upper bound is achieved for $F = \{1, 2, \dots, 2^{N-1}\}$. Let us, from here on, assume that $F = \{1, 2, \dots, 2^{N-1}\}$ and that $N \geq 2$.

By the above formula :

$$\delta(\mathcal{G}_P^n, \mathcal{M}_a)/2 \equiv P_\theta^n(\hat{\theta}_n \neq \theta) = \frac{N-1}{N^n} ; n=2, 3, \dots, .$$

It follows then that $\lim_{N \rightarrow \infty} \delta(\mathcal{G}_P^n, \mathcal{M}_a)/2 = 1$ or $= 0$ as $n=1$ or $n \geq 2$.

If only one observation X_1 is available, then θ is located to the N -point set $\{1-X_1, 2-X_1, \dots, 2^{N-1}-X_1\}$. If, however, another observation X_2 is available and $X_2 \neq X_1$, (this have probability $1 - \frac{1}{N}$), then θ is completely known. Thus we see that the phenomenon is related to the uniqueness of dyadic expansions.

Note also that $\sqrt[n]{\delta(\mathcal{G}_P^n, \mathcal{M}_a)/2}$ is strictly decreasing in n . In contradistinction to this we have seen, theorem 6.16, that $\sqrt[n]{\delta(\mathcal{G}^n, \mathcal{M}_a)/2} \rightarrow \sup_n \sqrt[n]{\delta(\mathcal{G}^n, \mathcal{M}_a)/2}$ for any dichotomy \mathcal{G} .

Furthermore

$$C([\mathcal{G}_P]_{\{0, \theta\}}) = \sum_x P(x)^{1-t} P(x-\theta)^t = \frac{1}{N} \text{ or } = 0 \text{ as } \theta \in F-F \text{ or not. Hence } C(\mathcal{G}) = \tau(\mathcal{G}) = \sigma(\mathcal{G}) = \frac{1}{N} .$$

10 References.

- [1] Blackwell, D. (1951). Comparison of experiments. Proc. Second Berkeley Sympos. Math.Statist.Probab. 93-102.
- [2] Blackwell, D. (1953). Equivalent comparisons of experiments. Ann.Math.Statist. 24, 265-272.
- [3] Blackwell, D. and Girshick, M.A. (1954). Theory of Games and Statistical Decisions. John Wiley & Sons.
- [4] Bohnenblust, F. Shapley, L. Shermann, S. (1951). Unpublished Rand Memorandum.
- [5] Boll, C. (1955). Comparison of experiments in the infinite case. Ph.D. Thesis. Stanford University.
- [6] Chernoff, H. (1952). A measure of asymptotique effeciency for tests of a hypothesis based on the sum of observations. AMS, 23, 493-507.
- [7] De Groot, M.H. (1962). Uncertainty, information and sequential experiments. Ann.Math.Statist. 33 , 404-419.
- [8] Dvoretzky, A. Wald, and Wolfowitz, J. (1951). Elimination of randomization in certain statistical decision procedures and zero-sum two person games. Ann.Math.Stat. 22, 1-21.
- [9] Efron and Truax. (1968) Deviation theory in exponential families. AMS 39, 1402-1424.
- [10] Heyer, H. (1969). Erschöpftheit und Invarianz beim vergleich von experimenten. Z. Wahrscheinlichketstheorie verw. Geb. 12, 21-55.
- [11] Heyer, H. (1973). Mathematische Theorie statistischer Experimente. Springer-Verlag.
- [12] Le Cam. L. (1964). Sufficiency and approximate sufficiency. Ann.Math.Statist. 35, 1419-1455.
- [13] Le Cam. L. (1968). Théorie asymptotique de la décision statistique. Les presses de l'Université de Montreal.
- [14] Le Cam, L. (1974). On the information contained in additional observations. Ann.Statist. 2, 630-649.

- [15] Le Cam, L. (1974). Notes on asymptotic methods in statistical decision theory. Centre de Recherches Math. Univ. de Montréal.
- [16] Lindqvist, B. (1976) How fast does a Markov chain forget the initial state ? A decision theoretic approach. Statistical research report. Institute of mathematics, Univ. of Oslo. No. 2, 1976.
- [17] Lyapunov, A. (1940). Sur les fonctions-vecteurs complètement additives. Izvestia Akad.Nauk. SSR. Ser.Mat. 4, 465-478.
- [18] Morse, N. and Sacksteder, R. (1966). Statistical isomorphism. Ann.Math.Statist. 37, 203-214.
- [19] Torgersen, E.N. (1970). Comparison of experiments when the parameter space is finite. Z. Wahrscheinlichkeits-theorie Verw. Geb. 16, 219-249.
- [20] Torgersen, E.N. (1972). Comparison of translation experiments. Ann.Math.Stat. 43, 1383-1399.
- Torgersen, E.N. Statistical research reports. Dep. of Math. Univ. of Oslo :
- [21] No. 3, (1974). Comparison of experiments by factorization.
- [22] No. 6, (1974). Asymptotic behaviour of powers of dichotomies.
- [23] Torgersen, E.N. (1975). Notes on comparison of statistical experiments. Chapters 0-8. Recorded and supplemented by Bo Lindqvist. Statistical memoir. Dep. of Math. Univ. of Oslo. No. 1, 1975.

Corrections to : DEVIATIONS FROM TOTAL INFORMATION AND FROM TOTAL
IGNORANCE AS MEASURES OF INFORMATION. Statistical Research Report
No. 3, 1976, by Erik N. Torgersen

page	line	old version	new version
3.3	10 ↓	$P_{\theta}(A)$	$P_{\theta}(A_{\theta})$
3.5	7 ↓	distribution $ \theta\rangle$.	distribution on $\theta \mathcal{G}$.
7.9	1 ↓	\tilde{K}_n	K_n
8.2	4 ↑	$y_i \in F\}$.	$Y_i \in F\}$.
8.3	2 ↓	$[r(\xi)\eta]$	$[r(\xi)\eta^5]$
"	2 ↑	"	"
"	1 ↑	"	"
8.5	4 ↑	"	"
"	2 ↑	"	"
8.6	3 ↑	$\delta(\mathcal{G}_{\{\theta_1, \theta_2\}}^n)$	$\delta(\mathcal{G}_{\{\theta_1, \theta_2\}}^n, \mathcal{M}_a)$
8.9	5 ↑	of θ	of estimators of θ
8.12	9 ↑	σ -finite measure	probability measure
"	6 ↑	bounded	μ -integrable.
8.19	11 ↑	$t_r(\mathcal{G}^n)$	$\kappa_r(\mathcal{G}^n)$
8.20	8 ↑	max	sup
"	"	min	inf
"	7 ↑	max	sup
9.7	5 ↓	estimator estimator	estimator